Supplementary Materials

Habitat heterogeneity and phenotypic variation: Site temperature largely predicts diverse spawning portfolios of an interior Chinook Salmon stock

07 October, 2025

Contents

S1 Datasets

S1.1 Spawn timing data

Spawn timing data for Chinook salmon were collected from 2001 to 2005 in the MFSR. Because redds were not observed daily, we inferred spawn dates as the initial date (day of year) a completed redd was observed.

We removed data from 2001, and data from Knapp Creek and Cape Horn Creek, as these sites were not consistently sampled.

We spatially joined each redd GPS location to the NHDPlus Version 2 (Horizon Systems, 2018) to assign stream reaches based on a common identifier (COMID). The COMID is used to link redd data with covariate data associated with the stream reach on which it is located. A summary of the dataset is shown in Table S1.1.

Table S1.1: First five rows of the Chinook salmon redd dataset.

redd_id	COMID	spawn_date	stream	year	yday
1295	23519365	2002-08-23	Bear Valley	2002	235
1296	23519365	2002-08-23	Bear Valley	2002	235
1298	23519319	2002-08-23	Bear Valley	2002	235
1299	23519319	2002-08-23	Bear Valley	2002	235
1300	23519319	2002-08-23	Bear Valley	2002	235

S1.1.1 Evaluate Grouping structure

The structure of the data is repeated measures on COMIDs, and multiple years, shared across COMIDs within streams.

Many `COMIDs` only have 1-2 observation (Figure S1.1), so the groups are not well sampled. There are 23 COMIDs with <5 redds (26%), 13 with <= 2. (12.5%). With <5 obs/level, variance estimates can become unstable and lead to overfitting and absorbing noise (low AIC / high R2) and singular fits.

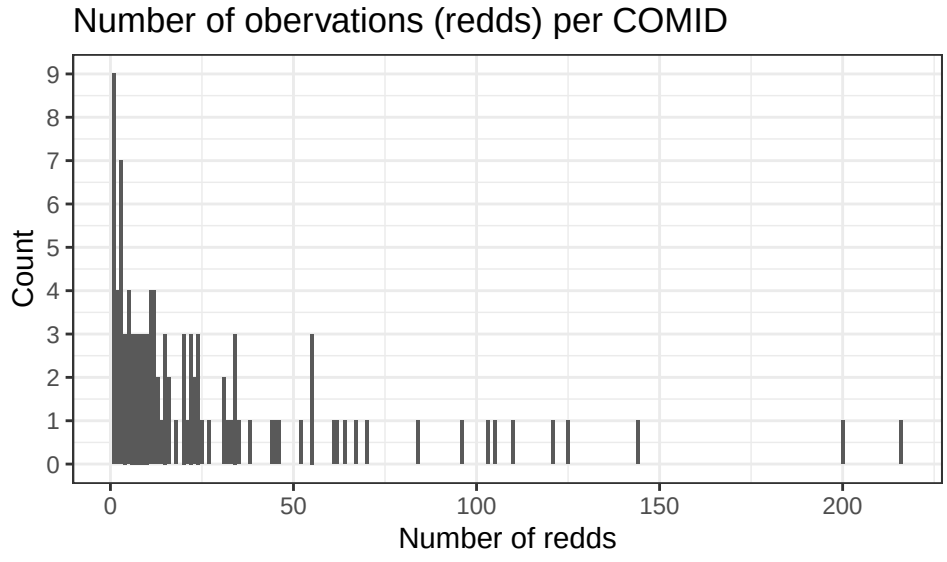


Figure S1.1: Number of observations (redds) per COMID.

S1.1.2 Spawn data descriptive statistics and visualizations

Variation by stream, year, and COMID (stream reach)

Figure S1.2 provides a visual summary of the distribution of spawn timing (day of year, `yday`), and variation by stream system and year.

Spawn timing is generally unimodal, with a peak in late August to early September. The distribution is slightly left skewed, but the mean and median are similar, indicating low skewness. The variance is low, suggesting no overdispersion. Suggests Poisson family for response if count of spawning events per day, or Gaussian if modeling day of year as continuous.

There is also substantial variation in spawn date by stream and year (Fig. S1.2B-C; Fig. S1.3) and within COMIDs within streams (Fig. S1.4).

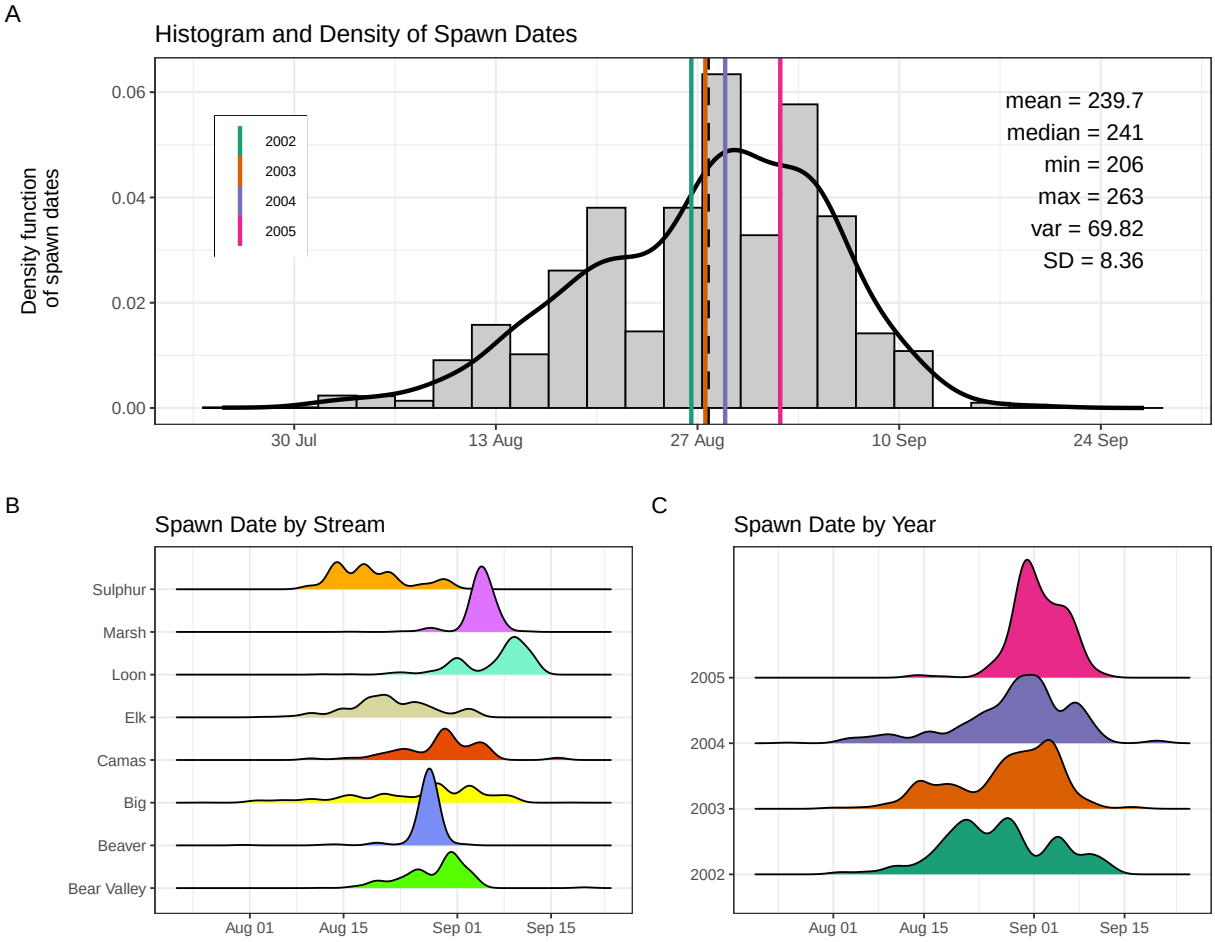


Figure S1.2: Histogram and density of spawn timing for all streams and years (A), and by stream (B) and year (C). In panel (A), colored, vertical lines indicate the mean spawn date for each year., and the black vertical line indicated the global mean.

Cumulative proportion of redds over time

The temporal progression of Chinook salmon spawning activity varied within each stream across the four study years (Figure S1.5). The cumulative proportion of redds provides an intuitive measure of the spawning season's pace and timing. Streams like Marsh and Sulphur exhibit rapid increases, suggesting short,

concentrated spawning windows. In contrast, streams like Big and Camas show more gradual increases, indicating a more protracted spawning season. Year-to-year variation is evident in several streams. Overall, this figure highlights both spatial and temporal heterogeneity in spawn timing that underpins the need for models incorporating stream- and year-specific effects.

```
## Converting page 1 to C:/Users/BryanMaitland/Projects/mfsr_phenology/plots/Figure_03.png... done!
```

```
## [1] "C:/Users/BryanMaitland/Projects/mfsr_phenology/plots/Figure_03.png"
```

S1.2 Covariate Derivation and Screening

To identify environmental predictors of Chinook salmon spawn timing, we quantified covariates that describe thermal and physical stream conditions at redd locations. Covariates were selected based on three criteria: (1) demonstrated influence on salmon phenology in prior literature, (2) availability at all redd locations (i.e., COMIDs), and (3) low collinearity with other predictors.

The initial focal covariates included stream temperature (°C), stream discharge (cms), elevation (m), and stream slope (unitless). Elevation and slope were extracted from NHDPlus (Moore et al. 2019).

S1.2.1 Elevation and slope

Elevation and stream slope data were available at the COMID (stream reach) scale from the NHDPlus Version 2 (Horizon Systems, 2018).

S1.2.2 Stream temperature

We used modeled daily average stream temperature values predicted at the stream segment (COMID) scale (Siegel et al. 2023). These data were downloaded and filtered to 2002-2005 and for the MFSR. Figure S1.6 shows the modeled thermal regimes for MFSR tributaries.

These modeled temperatures were validated against empirical logger data collected at a subset of sites in the Middle Fork basin, with strong agreement ($R^2 > 0.9$).

ADD PLOT OF COMPARISON BETWEEN EMPIRICAL AND MODELED TEMPPS

S1.2.3 Discharge (streamflow)

Stream flow data were compiled from a single USGS Gage lower in the watershed (MF Salmon River at MF Lodge NR Yellow Pine ID - 13309220; Figure S1.7).

S1.2.4 Time-windowed temp and flow metrics

For each redd, we computed several time-windowed summaries of temperature and flow relative to the inferred spawn date:

- **Antecedent metrics**: average conditions over 30, 60, and 90 days preceding the spawn date (e.g., `temp_90_before`, `flow_60_before`),
- **Spanning metrics**: averages over time windows centered on the spawn date,
- **Post-spawn metrics**: 30-, 60-, and 90-day averages following the spawn date,
- **Time-invariant metrics**: summaries calculated relative to a fixed calendar date (e.g., August 1) representing an approximate onset of the spawning window.

The time invariant and after metrics were omitted from further consideration as preliminary data exploration showed weak if any relationship with spawn timing.

Table S2.1: AIC comparison of linear mixed models relating spawn date (‘yday‘) to ‘temp_90‘.

Model	df	AIC	delta
m.t90.3	5	15848.81	0.0000
m.t90.2	4	16570.12	721.3107
m.t90.1	3	18629.30	2780.4987

Table S2.2: Model performance of linear mixed models relating spawn date (‘yday‘) to ‘temp_90‘.

Name	Model	R2_conditional	R2_marginal	ICC	RMSE	AIC_wt	BIC_wt	Performance_Score
m.t90.3	lmerMod	0.9177407	0.7643117	0.6509825	3.108315	1	1	0.8333333
m.t90.2	lmerMod	0.9086448	0.6915320	0.7038423	3.489863	0	0	0.6100606
m.t90.1	lmerMod	0.6571743	0.0000000	0.6571743	4.929494	0	0	0.0195227

S2 Exploratory Data Analysis

S2.1 Bivariate relationships

We did exploratory data analysis on the compiled dataset to identify candidate covariates for inclusion in model selection.

Bivariate relationships between spawn date (`yday`) and continuous predictors (Figure S2.1), and pairwise correlations between all continuous covariates (Figure S2.2), are shown below.

Takeaways from Figures S2.1 **and** S2.2:

Amoung temperature variables, `temp_90` clearly shows the strongest relationship with spawn date (Figure S2.1), and is highly colinear with `temp_30` and `temp_60` (Figure S2.2).

The is a weak negative relationship between `mean_elevation` and `yday` (Figure S2.1), and it is weakly correlated with `temp_90` (0.31, Figure S2.2).

- `flow_30` = decaying exponential, obvious year effect
- `flow_60` = ditto
- `flow_90` = inflections, year effect clear
- `slope` = no relationship

S2.2 temp_90

Figure S2.3 shows the relationship between `temp_90` and spawn date by stream and year.

We evaluated the role of `temp_90` in explaining variation in Chinook salmon spawn timing:

```
m.t90.1 <- lmer(yday ~ (1|COMID), data = df_mod, REML = FALSE)
m.t90.2 <- lmer(yday ~ temp_90 + (1|COMID), data = df_mod, REML = FALSE)
m.t90.3 <- lmer(yday ~ temp_90 + I(temp_90^2) + (1|COMID), data = df_mod, REML = FALSE)
```

The model with a quadratic effect of `temp_90` (`m.t90.3`) improved dramatically over the null (intercept-only) model and linear effect (Table S2.1), and boosting both conditional and marginal R^2 (Table S2.2). This confirms a strong, nonlinear effect of pre-spawn temperature on spawn timing (Figure S2.4).

Next, we add in stream and year as main effects, and then interactions between `temp_90` and stream and year:

Table S2.3: AIC comparison of linear mixed models relating spawn date (‘yday‘) to ‘temp_90‘.

Model	df	AIC	delta
m.t90.8	18	12136.83	0.0000
m.t90.7	22	13038.01	901.1783
m.t90.6	15	13473.12	1336.2837
m.t90.5	8	13531.17	1394.3321
m.t90.4	12	15807.98	3671.1454
m.t90.3	5	15848.81	3711.9726

Table S2.4: Model performance of linear mixed models relating spawn date (‘yday‘) to ‘temp_90‘.

Name	Model	R2_conditional	R2_marginal	ICC	RMSE	AIC_wt	BIC_wt	Performance_Score
m.t90.8	lmerMod	0.9928189	0.6448836	0.9797782	1.581472	1	1	0.8333333
m.t90.7	lmerMod	0.9852284	0.7177821	0.9476590	1.865263	0	0	0.5267941
m.t90.6	lmerMod	0.9800601	0.7293886	0.9263154	2.022117	0	0	0.5061439
m.t90.5	lmerMod	0.9871140	0.6600787	0.9620912	2.021778	0	0	0.4528690
m.t90.3	lmerMod	0.9177407	0.7643117	0.6509825	3.108315	0	0	0.2196386
m.t90.4	lmerMod	0.8982595	0.7921275	0.5105629	3.110542	0	0	0.1666667

```
m.t90.4 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + (1|COMID), data = df_mod, REML = FALSE)
m.t90.5 <- lmer(yday ~ temp_90 + I(temp_90^2) + year + (1|COMID), data = df_mod, REML = FALSE)
m.t90.6 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + (1|COMID), data = df_mod, REML = FALSE)
m.t90.7 <- lmer(yday ~ temp_90 * stream + I(temp_90^2) + year + (1|COMID), data = df_mod, REML = FALSE)
m.t90.8 <- lmer(yday ~ temp_90 * year + I(temp_90^2) + stream + (1|COMID), data = df_mod, REML = FALSE)
```

Table S2.3 shows that adding stream (`m.t90.4`) or year (`m.t90.5`) individually improved model fit relative to `m.t90.3`, but year alone (`m.t90.5`) performed better than stream alone (`m.t90.4`).

Combining stream and year additively (`m.t90.6`) provided further improvement in AIC, along with the best marginal R2 and RMSE (Table S2.4). This suggests that both stream and year provide meaningful structure in explaining variation in spawn timing.

Adding the interaction between temperature and stream (`m.t90.7`) increased complexity, slightly lowered the marginal R2, and has high multicollinearity (VIF ~15).

Adding the interaction between temperature and year (`m.t90.8`) gave the lowest AIC overall but had the lowest marginal R^2 of all models tested. This means that while the model “fit” the data better on paper (AIC), it did so by overfitting or by capturing year-to-year idiosyncrasies that may not generalize well to new data (suspicious marginal effect, Figure S2.5).

The suspicious curvature in the marginal effect plot of `m.t90.8` (showing an unrealistic predicted relationship between `temp_90` and `yday`) confirms the risk of confounding or overfitting with interaction terms—particularly given that the random intercepts already capture substantial site-level variation.

Given the trade-offs, model `m.t90.6` emerged as the best compromise: it retained additive main effects of temperature, stream, and year, had high conditional and marginal R^2, moderate AIC, low RMSE, and avoided high multicollinearity.

S2.2.1 Summary

The exploratory analysis of temperature effects on spawn timing revealed a clear, nonlinear relationship between pre-spawn temperature and the day of year when spawning occurs. A quadratic term for temperature

Table S2.5: AIC comparison of linear mixed models relating spawn date ('yday') to 'mean_elevation'.

Model	df	AIC	delta
m.ele.5	14	18329.29	0.00000
m.ele.6	18	18371.62	42.33148
m.ele.3	11	18431.88	102.58771
m.ele.7	10	18435.12	105.82567
m.ele.4	7	18501.44	172.14764
m.ele.2	4	18611.27	281.97710
m.ele.1	3	18629.30	300.01337

consistently improved model fit over linear-only specifications. When adding stream and year as main effects, model performance improved substantially, with both factors accounting for additional variation in spawn timing. However, introducing interaction terms between temperature and either stream or year led to either inflated multicollinearity (VIFs > 10) or implausible model predictions, indicating potential overfitting. Although the temperature-by-year interaction model (`m.t90.8`) had the lowest AIC, it yielded a counterintuitive relationship between temperature and spawn timing. Therefore, the additive model with quadratic temperature and main effects of stream and year (`m.t90.6`) was selected as the preferred model, balancing goodness-of-fit with interpretability and biological realism.

S2.3 `mean_elevation`

Figure S2.6 shows the relationship between between (A) `mean_elevation` and `temp_90` by stream and (B) `mean_elevation` and `yday` (spawn date) by stream and year.

We evaluated the role of mean elevation in explaining variation in Chinook salmon spawn timing:

```
m.ele.1 <- lmer(yday ~ (1|COMID), data = df_mod, REML = FALSE)
m.ele.2 <- lmer(yday ~ mean_elevation + (1|COMID), data = df_mod, REML = FALSE)
m.ele.3 <- lmer(yday ~ mean_elevation + stream + (1|COMID), data = df_mod, REML = FALSE)
m.ele.4 <- lmer(yday ~ mean_elevation + year + (1|COMID), data = df_mod, REML = FALSE)
m.ele.5 <- lmer(yday ~ mean_elevation + stream + year + (1|COMID), data = df_mod, REML = FALSE)
m.ele.6 <- lmer(yday ~ mean_elevation * stream + (1|COMID), data = df_mod, REML = FALSE)
m.ele.7 <- lmer(yday ~ mean_elevation * year + (1|COMID), data = df_mod, REML = FALSE)
```

Including mean elevation as a main effect improved model fit (ΔAIC = 19 between `m.ele.1` and `m.ele.2`). The best-supported model (`m.ele.5`) included mean elevation alongside stream and year (ΔAIC = 281 between `m.ele.2` and `m.ele.5`), with moderate collinearity (VIFs ~6), suggesting that mean elevation can contribute to explaining spawn timing when controlling for stream and year (S2.7).

However, adding interactions with stream or year substantially increased model complexity and collinearity, as evidenced by extremely high VIFs (>45). Given the known inverse relationship between elevation and temperature across streams (Figure S2.6A), caution is warranted when later incorporating both elevation and temperature in the same model.

S2.4 `slope`

`slope` is not related to `yday` (Figure S2.8A), though there is some association between `slope` and `yday` when considering stream (Figure S2.8B). We will likely drop slope.

Table S2.6: Model performance of linear mixed models relating spawn date ('yday') to 'mean_elevation'.

Name	Model	R2_conditional	R2_marginal	ICC	RMSE	Sigma	AIC_wt	AICc_wt	BIC_w
m.ele.5	lmerMod	0.6399737	0.5939968	0.1132428	4.876512	4.927506	1	1	
m.ele.7	lmerMod	0.6684326	0.1243275	0.6213568	4.774852	4.851946	0	0	
m.ele.4	lmerMod	0.6574860	0.1155940	0.6127186	4.835441	4.913296	0	0	
m.ele.1	lmerMod	0.6571743	0.0000000	0.6571743	4.929494	5.009972	0	0	
m.ele.2	lmerMod	0.6503692	0.1014398	0.6108988	4.929830	5.009158	0	0	
m.ele.6	lmerMod	0.6398941	0.6301923	0.0262347	4.989054	5.016809	0	0	
m.ele.3	lmerMod	0.6269782	0.5838335	0.1036716	4.971609	5.022058	0	0	

Table S2.7: Variance inflation factors for model 'm.ele.5'.

	GVIF	Df	GVIF^(1/(2*Df))
mean_elevation	5.739424	1	2.395709
stream	5.943076	7	1.135760
year	1.042077	3	1.006893

S2.5 `flow_90`

Because flow data are not COMID- or stream-specific, it makes sense to think about and represent flow as an out-of-basin year effect that determines when adults make it back to the MFSR and initially onto the spawning grounds.

The strong correlation between `flow_90` and `year` can be seen clearly in (Figure S2.9A).

Next we compare the following simple linear models to examine functional structure:

```
# flow_90 models
m1 <- lm(yday ~ flow_90, data = model_data)
m2 <- lm(yday ~ I(flow_90^2), data = model_data)
m3 <- lm(yday ~ flow_90 + year, data = model_data)
m4 <- lm(yday ~ flow_90 + year + stream, data = model_data)
m5 <- lm(yday ~ flow_90 * stream, data = model_data)
m6 <- lm(yday ~ flow_90 * stream + year, data = model_data)
m7 <- lm(yday ~ flow_90 * year + stream, data = model_data)
m8 <- lm(yday ~ flow_90 * stream + I(flow_90^2), data = model_data)
m9 <- lm(yday ~ flow_90 * stream + year + I(flow_90^2), data = model_data)
```

AIC scores suggest `m7` is best model (Table S2.9). However, while the R^2 value is 0.98, the parameter estimate for `flow_90` has is incredibly small and most variation is now attributed to the year contrasts (Table S2.10). Thus, `flow_90` is clearly confounded with `year`, and confirmed by high Variance Inflation Factor (VIF) scores (Figure S2.10).

Table S2.8: Variance inflation factors for model 'm.ele.6'.

	GVIF	Df	GVIF^(1/(2*Df))
mean_elevation	2.317261e+03	1	48.137937
stream	9.622172e+10	7	6.088629
mean_elevation:stream	5.226657e+11	7	6.870938

Table S2.9: Model performance of linear models relating spawn date (‘yday‘) to ‘flow_90‘.

Name	Model	R2	R2_adjusted	RMSE	Sigma	AIC_wt	AICc_wt	BIC_wt	Performance_Score
m7	lm	0.9780529	0.9779505	1.237710	1.240800	1	1	1	1.000000
m9	lm	0.9407030	0.9403269	2.034449	2.041229	0	0	0	0.496053
m6	lm	0.9108990	0.9103638	2.493858	2.501751	0	0	0	0.447289
m4	lm	0.8913935	0.8909958	2.753330	2.758824	0	0	0	0.418193
m3	lm	0.8745347	0.8743681	2.959322	2.961778	0	0	0	0.394284
m8	lm	0.7348812	0.7334668	4.301804	4.313980	0	0	0	0.217413
m5	lm	0.7130109	0.7115760	4.475722	4.487641	0	0	0	0.192156
m1	lm	0.5900974	0.5899614	5.348977	5.350752	0	0	0	0.057622
m2	lm	0.5352432	0.5350890	5.695650	5.697539	0	0	0	0.000000

Table S2.10: Parameter estimates for ‘m7‘, ‘yday flow_90 * year + stream‘.

Parameter	Coefficient	SE	CI	CI_low	CI_high	t	df_error	p
(Intercept)	278.2566486	0.2130585	0.95	277.8388932	278.6744041	1306.0106434	3001	0.0000000
flow_90	-0.0219390	0.0001136	0.95	-0.0221617	-0.0217163	-193.1695196	3001	0.0000000
year2003	-8.3607844	0.2371172	0.95	-8.8257131	-7.8958557	-35.2601350	3001	0.0000000
year2004	11.0240656	0.3856186	0.95	10.2679620	11.7801692	28.5879998	3001	0.0000000
year2005	1.0123700	0.7619753	0.95	-0.4816766	2.5064167	1.3286127	3001	0.1840768
streamBeaver	0.5591186	0.1048594	0.95	0.3535151	0.7647220	5.3320808	3001	0.0000001
streamBig	-0.4395848	0.0772852	0.95	-0.5911220	-0.2880475	-5.6878290	3001	0.0000000
streamCamas	-0.0934454	0.0899470	0.95	-0.2698094	0.0829186	-1.0388944	3001	0.2989375
streamElk	0.2219142	0.0861007	0.95	0.0530918	0.3907365	2.5773800	3001	0.0100026
streamLoon	0.0895360	0.1073905	0.95	-0.1210304	0.3001024	0.8337428	3001	0.4044923
streamMarsh	-0.7143977	0.0913025	0.95	-0.8934196	-0.5353759	-7.8245141	3001	0.0000000
streamSulphur	-0.0665471	0.1017872	0.95	-0.2661269	0.1330327	-0.6537861	3001	0.5132997
flow_90:year2003	0.0094506	0.0001190	0.95	0.0092172	0.0096840	79.3900831	3001	0.0000000
flow_90:year2004	-0.0101387	0.0002436	0.95	-0.0106163	-0.0096610	-41.6197601	3001	0.0000000
flow_90:year2005	-0.0046105	0.0005766	0.95	-0.0057411	-0.0034799	-7.9957348	3001	0.0000000

S2.5.1 Why `flow_90` is problematic

Not spatially resolved

- We are modeling spawn timing at the redd level (COMID/stream)
- But `flow_90` is calculated from a single downstream gauge, and applied to all redds
- This assumes flow conditions are identical across all sites
- Including it gives the illusion of spatially resolved variation that isn't there

Correlated with year

- Since `flow_90` varies mostly across years, it is strongly confounded with year
- Any flow-related signal is probably already captured by your year fixed effect
- Including both `flow_90` and year risks collinearity, and may produce misleading inferences

Spawn-time aligned flow experienced flow

- While `flow_90` is aligned to each redd's spawn date, it still reflects a lower-basin gauge, not the actual hydrologic conditions experienced at the redd site
- So it might be precisely wrong — aligned in time but irrelevant in space

Recommendation:

Drop `flow_90` from model.

Although we initially considered including 90-day mean streamflow (`flow_90`) as a predictor of spawn timing, this variable was ultimately excluded due to concerns about ecological validity and model overfitting. Stream flow data were derived from a single downstream USGS gauge and did not capture spatial variation across the study streams or reaches. Moreover, because `flow_90` was closely aligned with year, it introduced strong collinearity with the year effect and risked attributing site-level variation to flow patterns not actually experienced by individual redds. As such, we excluded `flow_90` to avoid misleading inference.

S3 Model fitting

The final dataset includes:

- `yday`: spawn date, continuous response variable
- `comid`, `stream`, `year`: grouping variables
- `temp_90`: 90-day mean temperature pre-spawn, continuous predictor variable
- `slope` and `mean_elevation`: provisionally retaining continuous predictor variable

We scaled the continuous covariates to have a mean of 0 and standard deviation of 1. This is important for mixed models, as it helps with convergence and interpretation (Table S3.1).

S3.1 Model specification

Table S3.1: Final dataset for modeling, first 5 rows.

yday	COMID	stream	year	temp_90	mean_elevation	slope
235	23519365	Bear Valley	2002	-0.3178564	0.7181159	-0.5189141
235	23519365	Bear Valley	2002	-0.3178564	0.7181159	-0.5189141
235	23519319	Bear Valley	2002	0.4665480	0.6954808	-0.7107062
235	23519319	Bear Valley	2002	0.4665480	0.6954808	-0.7107062
235	23519319	Bear Valley	2002	0.4665480	0.6954808	-0.7107062

Variable	Fixed?	Random?	Why
`COMID`	No	Yes	Not estimating COMID effects, just accounting for correlation, repeated measures
`Stream`	Yes	Maybe	8 streams to compare
`Year`	Yes	Maybe	4 years to compare

Here, `COMID` will be included in all models as a random (intercept) effect to account for repeated measures on `COMID`s. `Stream` will be treated as a fixed effect to compare average effects across streams, and `year` as a fixed effect to compare average effects across years. The later maybe be better included as a random effect, as it is not a treatment effect, but rather a random sample of years, but we only have 4 years of data, so we will treat it as a fixed effect for now.

S3.2 Model comparison

We used linear mixed-effects models to evaluate environmental predictors of Chinook salmon spawn timing, with redd observation day-of-year (`yday`) as the response variable. First, we scaled the continuous covariates to have a mean of 0 and standard deviation of 1 to assist convergence and interpretation.

We fit 31 additive models representing all combinations of fixed effects: `temp_90`, `stream`, `year`, `slope`, and `mean_elevation`. All models included a random intercept for `COMID` to account for repeated measures across stream reaches. Further, based on exploratory analysis and biological expectations of nonlinear thermal responses, temperature effects were modeled using both linear and quadratic terms for the 90-day average stream temperature prior to spawning (`temp_90` and `I(temp_90^2)`). Model comparison was performed using AIC, and model fit was evaluated using marginal and conditional R^2, RMSE, and intraclass correlation coefficients (ICC). All models were fit using maximum likelihood (REML = FALSE) to enable consistent comparison of differing fixed effects.

```
m1 <- lmer(yday ~ temp_90 + I(temp_90^2) + (1|COMID), data = df_mod, REML = FALSE)
m2 <- lmer(yday ~ stream + (1|COMID), data = df_mod, REML = FALSE)
m3 <- lmer(yday ~ year + (1|COMID), data = df_mod, REML = FALSE)
m4 <- lmer(yday ~ mean_elevation + (1|COMID), data = df_mod, REML = FALSE)
m5 <- lmer(yday ~ slope + (1|COMID), data = df_mod, REML = FALSE)

m6  <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + (1|COMID), data = df_mod, REML = FALSE)
m7  <- lmer(yday ~ temp_90 + I(temp_90^2) + year + (1|COMID), data = df_mod, REML = FALSE)
m8  <- lmer(yday ~ temp_90 + I(temp_90^2) + mean_elevation + (1|COMID), data = df_mod, REML = FALSE)
m9  <- lmer(yday ~ temp_90 + I(temp_90^2) + slope + (1|COMID), data = df_mod, REML = FALSE)
m10 <- lmer(yday ~ stream + year + (1|COMID), data = df_mod, REML = FALSE)
m11 <- lmer(yday ~ stream + mean_elevation + (1|COMID), data = df_mod, REML = FALSE)
m12 <- lmer(yday ~ stream + slope + (1|COMID), data = df_mod, REML = FALSE)
m13 <- lmer(yday ~ year + mean_elevation + (1|COMID), data = df_mod, REML = FALSE)
```

Table S3.3: AIC selection performance metrics for additive linear models; top 10 of 31 shown.

Model	df	AIC	delta_AIC	R2_marginal	R2_conditional	RMSE	ICC
m26	16	13356.04	0.000	0.785	0.956	2.023	0.796
m31	17	13358.04	1.998	0.785	0.956	2.023	0.795
m27	16	13468.96	112.925	0.737	0.979	2.022	0.922
m16	15	13473.12	117.079	0.729	0.980	2.022	0.926
m29	10	13487.82	131.783	0.727	0.984	2.022	0.942
m19	9	13488.73	132.690	0.726	0.984	2.022	0.943
m7	8	13531.17	175.128	0.660	0.987	2.022	0.962
m20	9	13532.90	176.858	0.660	0.987	2.022	0.962
m17	13	15787.25	2431.210	0.783	0.880	3.112	0.446
m28	14	15789.04	2433.003	0.783	0.880	3.112	0.445

```
m14 <- lmer(yday ~ year + slope + (1|COMID), data = df_mod, REML = FALSE)
m15 <- lmer(yday ~ mean_elevation + slope + (1|COMID), data = df_mod, REML = FALSE)

m16 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + (1|COMID), data = df_mod, REML = FALSE)
m17 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + mean_elevation + (1|COMID), data = df_mod, REML = F
m18 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + slope + (1|COMID), data = df_mod, REML = FALSE)
m19 <- lmer(yday ~ temp_90 + I(temp_90^2) + year + mean_elevation + (1|COMID), data = df_mod, REML = FA
m20 <- lmer(yday ~ temp_90 + I(temp_90^2) + year + slope + (1|COMID), data = df_mod, REML = FALSE)
m21 <- lmer(yday ~ temp_90 + I(temp_90^2) + mean_elevation + slope + (1|COMID), data = df_mod, REML = FA
m22 <- lmer(yday ~ stream + year + mean_elevation + (1|COMID), data = df_mod, REML = FALSE)
m23 <- lmer(yday ~ stream + year + slope + (1|COMID), data = df_mod, REML = FALSE)
m24 <- lmer(yday ~ stream + mean_elevation + slope + (1|COMID), data = df_mod, REML = FALSE)
m25 <- lmer(yday ~ year + mean_elevation + slope + (1|COMID), data = df_mod, REML = FALSE)

m26 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + mean_elevation + (1|COMID), data = df_mod, R
m27 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + slope + (1|COMID), data = df_mod, REML = FA
m28 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + mean_elevation + slope + (1|COMID), data = df_mod,
m29 <- lmer(yday ~ temp_90 + I(temp_90^2) + year + mean_elevation + slope + (1|COMID), data = df_mod, RE
m30 <- lmer(yday ~ stream + I(temp_90^2) + year + mean_elevation + slope + (1|COMID), data = df_mod, REM

m31 <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + mean_elevation + slope + (1|COMID), data = d

m26a <- lmer(yday ~ temp_90 + I(temp_90^2) + year  + stream * mean_elevation + (1|COMID), data = df_mod
```

Based on AIC, the best-fitting model was `m26`, though its improvement in AIC and performance over `m31`, the full model which includes `slope`, was modest (Table S3.3. Suggests little benefit to including `slope` as a predictor.

`m27`, which excludes `mean_elevation`, had substantially worse AIC and performance than `m26` and `m31`, indicating that elevation contributes meaningfully to explaining spawn timing variation.

Model `m16`, which excludes both `mean_elevation` and `slope`, had nearly identical predictive accuracy (RMSE) and higher conditional R^2 compared to `m26` and `m31`, despite a slightly lower (0.73 vs 0.78) marginal R^2. The 117-point AIC difference between `m16` and `m26` likely reflects subtle improvements in likelihood fit due to topographic covariates.

Visual inspection of partial (model-based) relationships (Figure S3.1 B)) revealed that the modeled elevation effect was inconsistent with ecological expectations.

In this case, `m26` is picking up a positive association: after controlling for temperature and stream identity,

higher-elevation reaches are predicted to spawn later. This is contrary to the observed negative relationships between elevation, temperature, and spawn timing in several streams (Figure S2.6).

This could reflect either (a) a real but subtle ecological effect, or (b) statistical artifact from collinearity/confounding.

a) There are some reasons a residual positive elevation effect might make sense:

- Hydrology and snowmelt: At higher elevations, cooler conditions might delay snowmelt-driven flow cues, potentially pushing back spawning independent of temperature.
- Population differences: If higher-elevation reaches host subpopulations adapted to shorter growing seasons, they might time spawning later to avoid fry emerging into very harsh winter conditions.
- Migration lags: Reaches further upstream (which often correspond to higher elevations) could simply take longer for fish to reach, even if temperatures are similar.

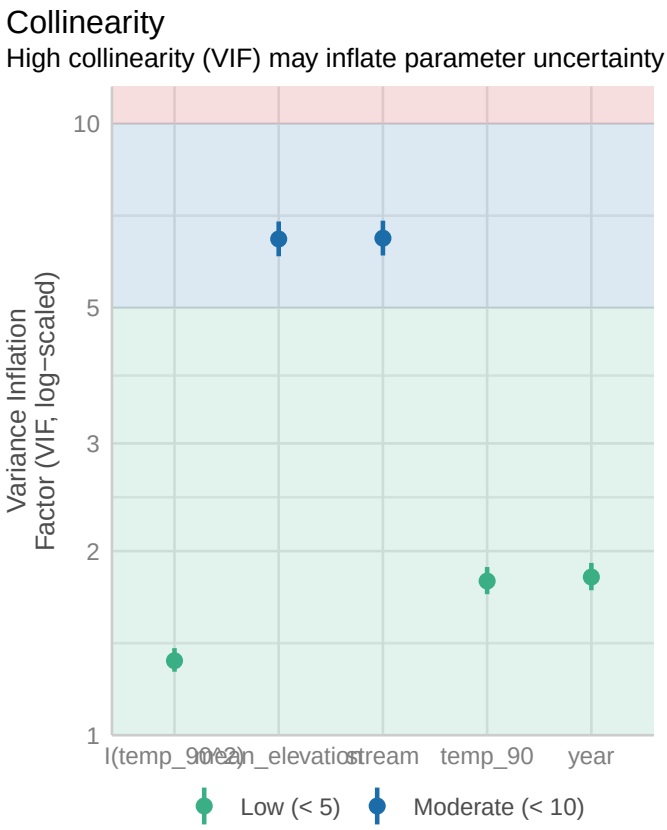
But these are second-order hypotheses—we'd need strong ecological justification to lean on them.

b) Statistical red flags

- The positive effect could just as easily be an artifact of collinearity. Elevation and temperature are tightly coupled in your dataset. Once you partial out temperature, the remaining variation in elevation is limited and unstable.
- The fact that predicted vs. observed plots showed inconsistent elevation effects across streams (your Figure elev-plots) suggests this may not be a robust, generalizable signal.

Variance inflation factors (VIFs) from m26 indicated moderate collinearity between elevation and stream (VIF = 6.50 and 6.48, respectively). Elevation was strongly confounded with stream identity in our dataset (Figure S2.6), with high-elevation streams showing considerable overlap in spawn timing and temperature.

```
##                    GVIF Df GVIF^(1/(2*Df))
## temp_90        1.786743  1        1.336691
## I(temp_90^2)   1.323720  1        1.150530
## stream         6.496841  7        1.143010
## year           1.814029  3        1.104352
## mean_elevation 6.477001  1        2.544995
```



I suggest we acknowledge both interpretations:

- Statistically: The positive elevation effect emerges after controlling for temperature, but it is likely confounded by collinearity.
- Ecologically: A true positive effect could be consistent with delayed migration, snowmelt-driven cues, or subpopulation adaptations, but our dataset can't fully disentangle these mechanisms.
- Conclusion: Because the elevation effect runs counter to raw data patterns and shows signs of confounding, we made the conservative decision to omit it from the final model set.

S3.3 Targeted model comparison

We compared our top additive model (`m16`; fixed effects: `temp_90`, `I(temp_90^2)`, `stream`, `year`; random intercept for `COMID`) to a series of increasingly flexible models to evaluate the contribution of random slopes and temperature nonlinearity.

S3.3.1 Random slopes for `temp_90`

We then tested whether model performance improved by allowing the effect of temperature to vary across stream reaches by adding a random slope for `temp_90` to the `COMID` grouping factor. This model was fit using restricted maximum likelihood (REML), and its fit was compared to the simpler random intercept model.

To account for potential variation in temperature sensitivity across sites, we extended `m16` to include **COMID-specific random slopes for `temp_90`**, yielding a model with the same fixed effects but with

Table S3.4: Model selection for 'm16' vs. 'm16_rs'.

Model	df	AIC	delta_AIC	R2_marginal	R2_conditional	RMSE	ICC
m16_rs	17	12948.76	0.000	0.698	0.985	1.781	0.949
m16	15	13459.05	510.283	0.714	0.981	2.022	0.932

Table S3.5: Model selection for 'm16_rs' vs. 'm16_rs_noquad'.

Model	df	AIC	delta_AIC	R2_marginal	R2_conditional	RMSE	ICC
m16_rs	17	12965.97	0.000	0.713	0.984	1.782	0.945
m16_rs_noquad	16	12986.41	20.441	0.714	0.984	1.788	0.943

the random structure `(1 + temp_90 | COMID)`. Because the random effects structure differed, we fit both models using restricted maximum likelihood (REML).

The random slope model (`m16_rs`) substantially improved fit (ΔAIC = 510), reduced RMSE (2.02 $\rightarrow$ 1.78 days), and slightly increased conditional R^2 (0.981 $\rightarrow$ 0.985). However, the marginal R^2 declined from 0.714 to 0.698, reflecting a redistribution of explanatory power from fixed to random effects. This shift is expected when site-specific variation in temperature responses is modeled explicitly, and suggests that temperature sensitivity varies meaningfully among stream reaches.

```
m16 <- lmer(
  yday ~ temp_90 + I(temp_90^2) + stream + year + (1 | COMID),
  data = df_mod, REML = TRUE
  )
m16_rs <- lmer(
  yday ~ temp_90 + I(temp_90^2) + stream + year + (1 + temp_90 | COMID),
  data = df_mod, REML = TRUE
  )
```

S3.3.2 Partitioning Variation Between Nonlinearity and Heterogeneity

To determine whether the **quadratic temperature term remained necessary** in the presence of random slopes, we re-fit the random slope model with and without `I(temp_90^2)` using maximum likelihood (ML).

The full model including both quadratic temperature and random slopes (`m16_rs`) outperformed the linear version (`m16_rs_noquad`) by ΔAIC = 20.4, despite only one additional degree of freedom. RMSE and R^2 values were nearly identical, but the full model retained a slightly better fit, particularly at the tails of the temperature distribution. This indicates that while **random slopes capture most of the temperature-related variation**, the quadratic term meaningfully refines the relationship without overfitting or destabilizing the model.

```
m16_rs <- lmer(
  yday ~ temp_90 + I(temp_90^2) + stream + year + (1 + temp_90 | COMID),
  data = df_mod, REML = FALSE
  )
m16_rs_noquad <- lmer(
  yday ~ temp_90 + stream + year + (1 + temp_90 | COMID),
  data = df_mod, REML = FALSE
  )
```

Table S3.6: Model selection for ‘m201‘ vs. ‘m202‘.

Model	df	AIC	delta_AIC	R2_marginal	R2_conditional	RMSE	ICC
m202	20	11568.63	0.000	0.620	0.994	1.359	0.985
m201	24	12948.51	1379.886	0.736	0.985	1.786	0.944
m16_rs	17	12965.97	1397.340	0.713	0.984	1.782	0.945

S3.3.3 Summary of random slopes and nonlinearity

For results

To account for site-level variation in thermal sensitivity, we extended this model by allowing COMID-specific random slopes for temperature. This significantly improved model fit (ΔAIC = 510), reduced prediction error (RMSE = 1.78 days), and increased conditional R^2, indicating that variation in temperature response among stream reaches was substantial and better captured as a random effect.

We then tested whether the quadratic temperature term remained necessary when random slopes were included. The full model with both random slopes and a quadratic temperature effect provided better support (ΔAIC = 20.4) than the linear version, suggesting that each component contributed complementary information. We retained this final model (`m16_rs`) as it provided strong predictive performance while acknowledging both general and site-specific patterns in temperature–spawn timing relationships.

S3.4 Interactions

To assess whether fixed-effect interactions provide additional explanatory power beyond what is captured by random slopes, we tested two models: one with a `temp_90 × stream` interaction (`m201`) and one with a `temp_90 × year` interaction (`m202`). Both were compared to the baseline random slope model with quadratic temperature (`m16_rs`).

```
m201 <- lmer(yday ~ temp_90 * stream + I(temp_90^2) + year + (1 + temp_90 | COMID), data = df_mod, REML
m202 <- lmer(yday ~ temp_90 * year + I(temp_90^2) + stream + (1 + temp_90 | COMID), data = df_mod, REML
```

Parameter estimates for interaction models.

m16_rs

m201: temp_90 * stream

m202: temp_90 * year

Predictors

Estimates

CI

p

Estimates

CI

p

Estimates

CI

p

(Intercept)

235.12

231.18 – 239.05

<0.001

234.55

230.27 – 238.83

<0.001

232.86

226.46 – 239.25

<0.001

temp 90

13.85

13.02 – 14.68

<0.001

14.99

13.22 – 16.76

<0.001

18.84

17.87 – 19.80

<0.001

temp 90^2

-0.86

-1.16 – -0.56

<0.001

-1.12

-1.43 – -0.82

<0.001

1.53

1.23 – 1.83

<0.001

stream [Beaver]

17.37

10.90 – 23.84

<0.001

17.93

11.04 – 24.83

<0.001

21.60

11.35 – 31.85

<0.001

stream [Big]

-0.17

-5.11 – 4.78

0.947

-1.07

-6.48 – 4.33

0.697

-3.78

-11.78 – 4.23

0.355

stream [Camas]

2.29

-2.84 – 7.42

0.381

2.00

-3.59 – 7.59

0.483

1.82

-6.51 – 10.14

0.668

stream [Elk]

11.88

5.12 – 18.64

0.001

11.74

4.53 – 18.95

0.001

15.89

5.13 – 26.64

0.004

stream [Loon]

2.85

-2.17 – 7.87

0.266

3.43

-2.21 – 9.06

0.233

1.71

-6.48 – 9.89

0.683

stream [Marsh]

7.91

2.55 – 13.27

0.004

9.15

3.35 – 14.96

0.002

9.50

0.82 – 18.18

0.032

stream [Sulphur]

14.69

8.97 – 20.42

<0.001

15.37

9.26 – 21.49

<0.001

24.50

15.44 – 33.56

<0.001

year [2003]

-4.99

-5.22 – -4.76

<0.001

-4.96

-5.19 – -4.73

<0.001

-7.03

-7.23 – -6.83

<0.001

year [2004]

2.76

2.52 – 3.00

<0.001

2.78

2.53 – 3.02

<0.001

2.96

2.77 – 3.15

<0.001

year [2005]

3.68

3.38 – 3.98

<0.001

3.69

3.39 – 4.00

<0.001

3.75

3.51 – 3.98

<0.001

temp 90 × stream [Beaver]

-2.81

-5.85 – 0.23

0.070

temp 90 × stream [Big]

0.83

-1.40 – 3.06

0.466

temp 90 × stream [Camas]

0.99

-1.42 – 3.39

0.421

temp 90 × stream [Elk]

0.68

-2.30 – 3.66

0.656

temp 90 × stream [Loon]

-1.29

-3.87 – 1.28

0.325

temp 90 × stream [Marsh]

-3.96

-6.48 – -1.44

0.002

temp 90 × stream[Sulphur]

-5.05

-7.75 – -2.35

<0.001

temp 90 × year [2003]

-3.89

-4.08 – -3.70

<0.001

temp 90 × year [2004]

0.63

0.40 – 0.87

<0.001

temp 90 × year [2005]

-1.29

-1.65 – -0.93

<0.001

Random Effects

2

3.36

3.37

1.96

00

45.06 COMID

50.42 COMID

114.83 COMID

11

12.18 COMID.temp_90

6.52 COMID.temp_90

17.85 COMID.temp_90

01

-0.23 COMID

-0.35 COMID

0.00 COMID

ICC

0.94

0.94

0.99

N

104 COMID

104 COMID

104 COMID

Observations

3016

3016

3016

Marginal R2 / Conditional R2

0.713 / 0.984

0.736 / 0.985

0.620 / 0.994

Model `m202` showed a large AIC improvement (ΔAIC -1397), but inspection of its predicted effects revealed implausible temperature relationships—specifically, an inverted quadratic curve inconsistent with biological expectations. This suggests overfitting or confounding in the interaction structure.

Model `m201` produced biologically reasonable predictions and modest AIC improvement (ΔAIC -18), but most interaction terms were non-significant, and variance explained (R^2) remained nearly unchanged. These results indicate that stream-specific variation in thermal sensitivity is already well-captured by the random slope structure.

Given the limited inferential gain and added complexity of the interaction models, we retained `m16_rs` (quadratic temperature with COMID-specific random slopes) as our final model.

S4 Model interpretation

The final model included a linear and quadratic effect of temperature (`temp_90`, `I(temp_90^2)`), additive fixed effects for `stream` and `year`, and a random intercept and slope for temperature by `COMID`:

Table S4.1: Parameter estimates for 'mod_final'.

Parameter	Coefficient	SE	CI	CI_low	CI_high	t	df_error	p	Effects	Group
(Intercept)	235.09	2.11	0.95	230.95	239.22	111.50	2999	0.00	fixed	
temp_90	13.90	0.43	0.95	13.06	14.74	32.50	2999	0.00	fixed	
I(temp_90^2)	-0.85	0.15	0.95	-1.15	-0.55	-5.54	2999	0.00	fixed	
streamBeaver	17.41	3.46	0.95	10.62	24.19	5.03	2999	0.00	fixed	
streamBig	-0.19	2.65	0.95	-5.37	5.00	-0.07	2999	0.94	fixed	
streamCamas	2.30	2.75	0.95	-3.08	7.69	0.84	2999	0.40	fixed	
streamElk	11.97	3.61	0.95	4.88	19.06	3.31	2999	0.00	fixed	
streamLoon	2.80	2.68	0.95	-2.47	8.06	1.04	2999	0.30	fixed	
streamMarsh	7.93	2.87	0.95	2.30	13.56	2.76	2999	0.01	fixed	
streamSulphur	14.69	3.06	0.95	8.69	20.69	4.80	2999	0.00	fixed	
year2003	-5.00	0.12	0.95	-5.23	-4.77	-42.79	2999	0.00	fixed	
year2004	2.77	0.12	0.95	2.53	3.01	22.57	2999	0.00	fixed	
year2005	3.68	0.15	0.95	3.38	3.98	24.00	2999	0.00	fixed	
SD (Intercept)	7.05	NA	0.95	NA	NA	NA	NA	NA	random	COMI
SD (temp_90)	3.55	NA	0.95	NA	NA	NA	NA	NA	random	COMI
Cor (Intercept~temp_90)	-0.23	NA	0.95	NA	NA	NA	NA	NA	random	COMI
SD (Observations)	1.83	NA	0.95	NA	NA	NA	NA	NA	random	Residua

```
mod_final <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + (1 + temp_90 | COMID),
  data = df_mod, REML = TRUE)
# mod_final <- lmer(yday ~ temp_90 + I(temp_90^2) + stream + year + mean_elevation + (1 + temp_90 | COM
#   data = df_mod, REML = TRUE)
# mod_final <- lmer(yday ~ temp_90 + I(temp_90^2) + year  + stream * mean_elevation + (1 + temp_90|COMI
```

This model balances explanatory power, biological realism, and parsimony. It captures both general patterns in spawn timing (via fixed effects) and local deviations in temperature response (via random slopes), and will serve as the basis for diagnostics, prediction, and ecological interpretation.

S4.1 Model summary and fit

Parameter estimates (Table S4.1) indicate a significant nonlinear effect of temperature on spawn timing, with the quadratic term (–0.85) confirming a concave-down relationship—i.e., spawn timing advances with increasing temperature, but levels off at high temperatures.

Most stream and year effects were significant, capturing expected spatiotemporal structure in spawn timing. Notably, Big, Camas, and Loon were not significantly different from the reference level (Bear Valley).

The random effects structure reveals substantial variation across COMIDs. The standard deviation for random intercepts ($\sqrt{\quad} = 7.05$) and random slopes for `temp_90` ($\sqrt{\quad} = 3.55$) indicates strong spatial heterogeneity in both baseline timing and temperature sensitivity (Table S4.1). The intraclass correlation (ICC) was high (0.95), reflecting strong grouping structure at the COMID level (Table S4.2).

Model fit was excellent: the marginal R^2 (variance explained by fixed effects) was 0.698, and the conditional R^2 (fixed + random effects) was 0.985 (Table S4.2). This suggests that the majority of explanatory power is derived from spatially varying thermal responses captured by the random slopes, with additional structure provided by the fixed effects.

Table S4.2: Model performance for 'mod_final'.

AIC	AICc	BIC	R2_conditional	R2_marginal	ICC	RMSE	Sigma
12948.76	12948.97	13050.96	0.9845502	0.6979933	0.9488428	1.780765	1.833613

S4.2 Residual diagnostics

Model diagnostics for the final model were generally strong (Figure S4.1). The posterior predictive check shows excellent agreement between the observed and model-predicted distributions of spawn timing, with overlapping density curves and no major deviations.

Plots of linearity and homoscedasticity indicate acceptable model performance. While the residual vs. fitted plot shows a slight trend, it does not appear strong enough to invalidate model assumptions. The spread of residuals is approximately constant across fitted values, though there is minor funneling at the lower end, likely reflecting skew in early spawn dates.

Influential observations are limited. A few data points exceed standard influence thresholds (|standardized residual| > 2 and high leverage), but none are extreme enough to warrant removal, and their leverage is modest. The model appears robust to outliers.

Collinearity is low: variance inflation factors (VIFs) for all fixed effects are well below the conservative threshold of 5, suggesting little concern about multicollinearity.

The normal Q-Q plot of residuals shows slight right-skew and some heavy tails, but the distribution is reasonably close to normal. Similarly, random effect Q-Q plots for intercepts and slopes show mostly linear trends with slight deviation at the tails, indicating acceptable assumptions for the mixed-effects structure.

Overall, the model shows no violations of key assumptions and is suitable for inference and prediction.

S4.3 Predictions against original data

Model predictions closely matched observed spawn timing, with predicted values aligning well along the 1:1 line (Figure S4.2). This supports strong overall model fit.

S4.4 Population-level effects

S4.4.1 Marginal means and contrasts of `yday` at each factor level

We estimated marginal mean of `yday` at each factor level, averaging over the random effects, to provide an overall estimate of the effect in the population (Figure S4.3).

Significant differences in spawn timing were observed between many stream pairs (Table S4.3), particularly involving Loon (later spawning) and Sulphur (earlier spawning). For example, fish in Loon spawned significantly later than in Bear Valley, Camas, and Elk, while Sulphur exhibited significantly earlier timing than all other streams except Elk. These patterns reflect spatial heterogeneity in temperature and elevation across streams that is not fully captured by fixed effects alone.

There was a clear trend toward later spawning over the four-year period (Table S4.4). Spawning in 2005 occurred significantly later than in all previous years. Differences between 2002 and 2003 were not statistically significant, but later years (2004 and especially 2005) were associated with a progressive delay in mean spawn timing. This temporal shift likely reflects interannual variability in temperature and flow conditions.

Table S4.3: Pairwise contrasts among stream effects on predicted spawn timing. Contrasts represent estimated differences in predicted spawn day of year between streams from the final model. Positive values indicate later predicted spawn timing in the first stream compared to the second. P-values are uncorrected.

Level1	Level2	Difference	SE	CI_low	CI_high	t	df	p
Beaver	Bear Valley	-1.43	3.47	-8.23	5.38	-0.41	2999	0.68
Big	Bear Valley	-2.66	2.66	-7.88	2.56	-1.00	2999	0.32
Camas	Bear Valley	0.63	2.75	-4.77	6.03	0.23	2999	0.82
Elk	Bear Valley	-6.60	3.62	-13.71	0.50	-1.82	2999	0.07
Loon	Bear Valley	8.90	2.68	3.64	14.15	3.32	2999	0.00
Marsh	Bear Valley	6.72	2.87	1.08	12.35	2.34	2999	0.02
Sulphur	Bear Valley	-9.47	3.16	-15.66	-3.28	-3.00	2999	0.00
Big	Beaver	-1.23	3.18	-7.46	4.99	-0.39	2999	0.70
Camas	Beaver	2.06	3.27	-4.35	8.46	0.63	2999	0.53
Elk	Beaver	-5.18	4.01	-13.04	2.69	-1.29	2999	0.20
Loon	Beaver	10.33	3.24	3.98	16.67	3.19	2999	0.00
Marsh	Beaver	8.14	3.41	1.46	14.82	2.39	2999	0.02
Sulphur	Beaver	-8.04	3.54	-14.98	-1.10	-2.27	2999	0.02
Camas	Big	3.29	2.40	-1.42	8.00	1.37	2999	0.17
Elk	Big	-3.94	3.35	-10.51	2.62	-1.18	2999	0.24
Loon	Big	11.56	2.35	6.95	16.16	4.92	2999	0.00
Marsh	Big	9.38	2.57	4.33	14.42	3.64	2999	0.00
Sulphur	Big	-6.81	2.78	-12.26	-1.36	-2.45	2999	0.01
Elk	Camas	-7.23	3.43	-13.97	-0.50	-2.11	2999	0.04
Loon	Camas	8.27	2.45	3.47	13.06	3.38	2999	0.00
Marsh	Camas	6.09	2.66	0.87	11.30	2.29	2999	0.02
Sulphur	Camas	-10.10	2.91	-15.80	-4.40	-3.47	2999	0.00
Loon	Elk	15.50	3.40	8.84	22.17	4.56	2999	0.00
Marsh	Elk	13.32	3.56	6.34	20.30	3.74	2999	0.00
Sulphur	Elk	-2.87	3.71	-10.14	4.41	-0.77	2999	0.44
Marsh	Loon	-2.18	2.57	-7.23	2.87	-0.85	2999	0.40
Sulphur	Loon	-18.37	2.91	-24.07	-12.66	-6.31	2999	0.00
Sulphur	Marsh	-16.19	3.10	-22.27	-10.10	-5.22	2999	0.00

Table S4.4: Pairwise contrasts among year effects on predicted spawn timing. Contrasts represent estimated differences in predicted spawn day of year between years from the final model. Positive values indicate later spawning in the first year compared to the second. P-values are uncorrected.

Level1	Level2	Difference	SE	CI_low	CI_high	t	df	p
2003	2002	0.97	0.57	-0.15	2.10	1.70	2999	0.09
2004	2002	2.35	0.65	1.07	3.62	3.61	2999	0.00
2005	2002	6.18	0.61	4.98	7.37	10.17	2999	0.00
2004	2003	1.37	0.61	0.18	2.56	2.26	2999	0.02
2005	2003	5.20	0.71	3.80	6.60	7.28	2999	0.00
2005	2004	3.83	0.87	2.11	5.54	4.38	2999	0.00

S4.4.2 Estimating response vs. relation

To visualize the model's predictions across the temperature gradient, we estimated the relationship between spawn timing and 90-day pre-spawn mean temperature (`temp_90`) for different groupings: overall, by stream, and by year. This approach allows us to assess both general trends and context-specific responses.

Stream-specific predictions (see plot suggestion below) show a consistent pattern: spawn timing increases nonlinearly with 90-day mean stream temperature, leveling off at high temperatures. This plateau is consistent with biological expectations, as spawning may be constrained by environmental or physiological thresholds. Stream-to-stream variation in predicted timing reflects both fixed stream effects and COMID-specific random intercepts and slopes.

Year-specific predictions similarly show consistent thermal responses across years, with modest offsets in average spawn timing due to year effects. These predictions further validate the model's generalizability and temporal consistency.

In addition, a combined stream-by-year plot (e.g., facet grid) confirms that the final model captures heterogeneity in both space and time without overfitting. Lines track the raw data closely across groups, particularly in mid-range temperatures where most observations are concentrated.

Together, these plots indicate that the final model effectively captures both nonlinear temperature effects and spatial variation in thermal sensitivity, while maintaining interpretability and predictive strength.

When stratified by stream and year, predictions captured both the nonlinear temperature response and variation in baseline spawn timing across sites and years (Figure S4.4). The curvature was consistent across years, while intercept shifts reflected known spatial patterns (e.g., later spawning in warmer, downstream reaches). These results confirm that the model accurately describes both average and context-specific phenological responses to temperature.

S4.5 Group-level effects (deviations from fixed effects)

Random intercepts and slopes varied considerably among COMIDs, reflecting spatial heterogeneity in both average spawn timing and thermal sensitivity (Figure S4.5A). We found considerable spread in intercepts, reflecting differences in average spawn timing between reaches. The random slopes for `temp_90` likewise varied meaningfully across COMIDs, indicating that temperature–spawn timing relationships are not constant across space.

Sites with earlier average spawn timing (lower intercepts) generally exhibited stronger positive responses to temperature (higher slopes), while later-spawning sites tended to show weaker temperature effects, a pattern also evident in the weak negative correlation between intercepts and slopes (r = -0.2; Figure S4.5B). This indicates that reaches with earlier average spawn timing (i.e., negative intercepts) tend to exhibit stronger temperature sensitivity (i.e., steeper positive slopes), whereas later-spawning reaches show weaker responses to temperature. Biologically, this suggests that early-spawning populations are more phenologically plastic and adjust their timing more closely to thermal cues, while late-spawning populations may be constrained by other factors—such as photoperiod, migration fatigue, or compressed spawning windows—resulting in diminished thermal responsiveness. These findings highlight spatial heterogeneity in phenological flexibility, with potential implications for how different populations may respond to climate change.

When grouped by stream, Bear Valley and Big Creek exhibited early average spawn timing and high thermal sensitivity, while Sulphur and Marsh Creeks had later average timing with flatter temperature responses (Figure S4.6).

However, when examining individual random effects for each COMID and stream, we observed considerable variation in both intercepts and slopes within streams (Figure S4.7). For example, Bear Valley and Big Creek had some of the earliest average spawn timings, but also exhibited a wide range of thermal sensitivities. In contrast, Sulphur Creek had later average spawn timing but also showed considerable variability in its response to temperature.

Variation in temperature–spawn timing relationships across stream reaches (COMIDs) as estimated from the final mixed-effects model is shown in Figure S4.8). While the overall relationship is positive and nonlinear, individual COMID slopes and intercepts vary considerably. Some reaches show steeper increases in spawn timing with temperature (i.e., stronger thermal sensitivity), while others are relatively flat, indicating a weaker or more buffered response. Grouping by stream shows that some streams (e.g., Bear Valley, Marsh) exhibit tightly clustered trajectories, while others (e.g., Camas, Big) show more divergence. This variation likely reflects fine-scale differences in local hydrology, geomorphology, or biological factors that influence how fish respond to thermal cues within stream systems. The consistency of the overall trend, despite local heterogeneity, supports the biological relevance of temperature in structuring spawn timing.

S4.6 Elevation effects embedded in random structure

Although mean elevation was excluded as a fixed effect due to collinearity with temperature and inconsistent global directionality, examination of random effects against elevation reveals spatial structure that elevation helps explain (Figure X). Random intercepts (average spawn timing) showed positive relationships with elevation in some streams (e.g., Big Creek), where higher-elevation sites tended to have later average spawning relative to the population mean. In contrast, other streams (e.g., Bear Valley, Beaver) showed little or no elevation pattern.

Random slopes (thermal sensitivity to temperature) exhibited similarly idiosyncratic patterns. In some cases (e.g., Camas, Marsh), thermal sensitivity declined with elevation, suggesting that fish at higher elevations may respond less strongly to interannual temperature variability. In others, relationships were weak or even opposite in direction.

Taken together, these patterns indicate that elevation influences both average spawn timing and thermal sensitivity, but in ways that differ across streams. This heterogeneity justifies our decision to capture elevation-linked variation through COMID-level random effects rather than imposing a single fixed elevation term.

S5 References

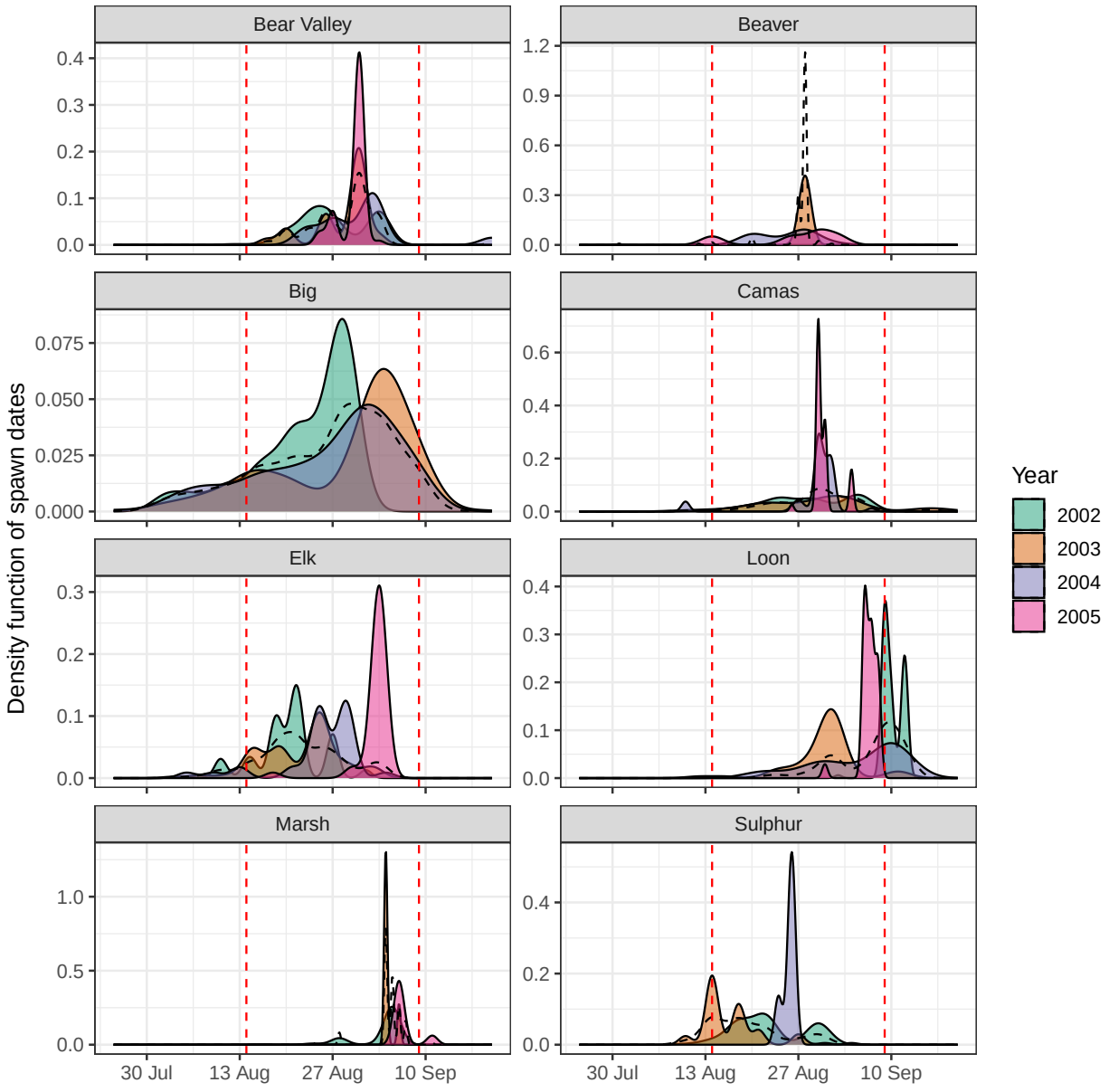


Figure S1.3: Spawning phenology of adult Chinook Salmon. In all panels, the black density function represented stream-level spawn timing, while the colored density functions represent the spawn timing of individual years. The dashed vertical purple lines represent the 5th and 95th percentiles of the basin-wide spawn timing.

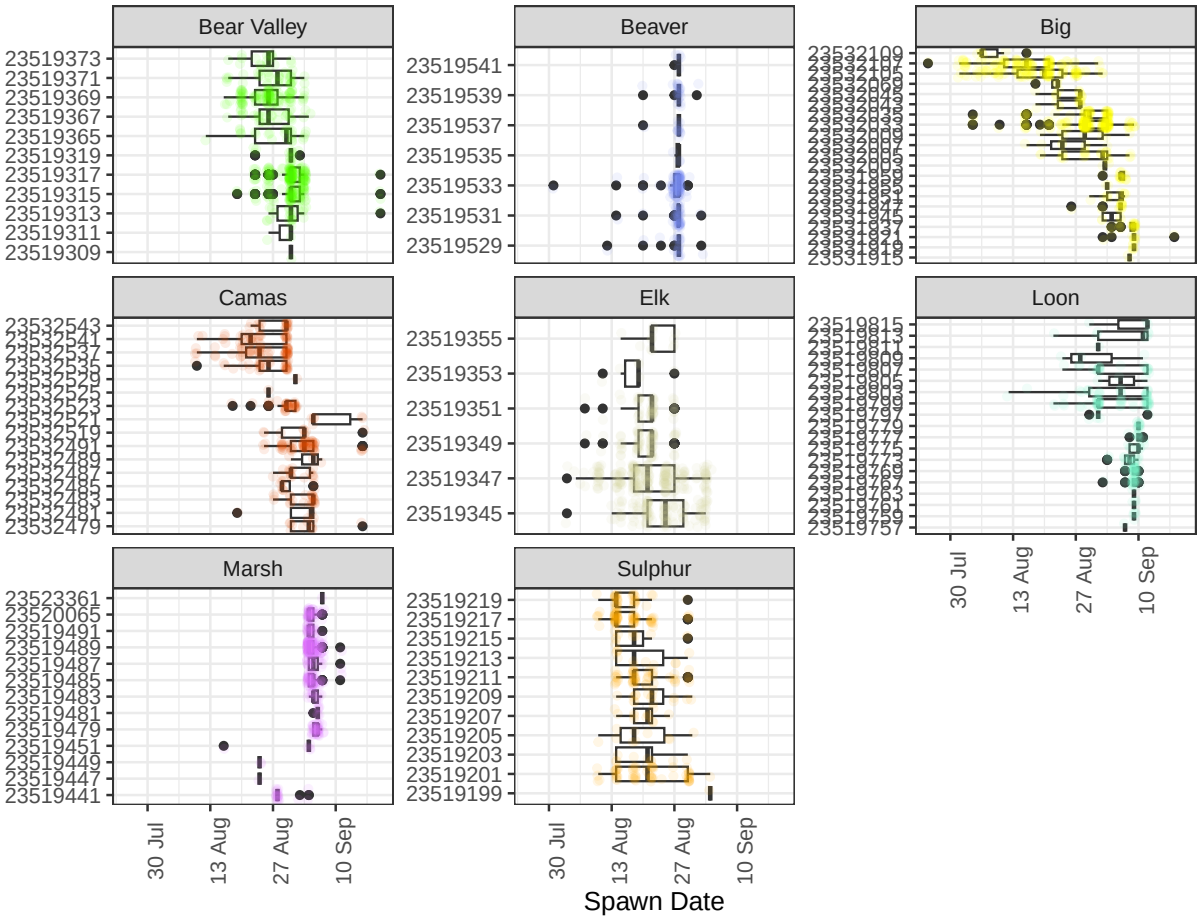


Figure S1.4: Spawn time variation by COMID and stream.

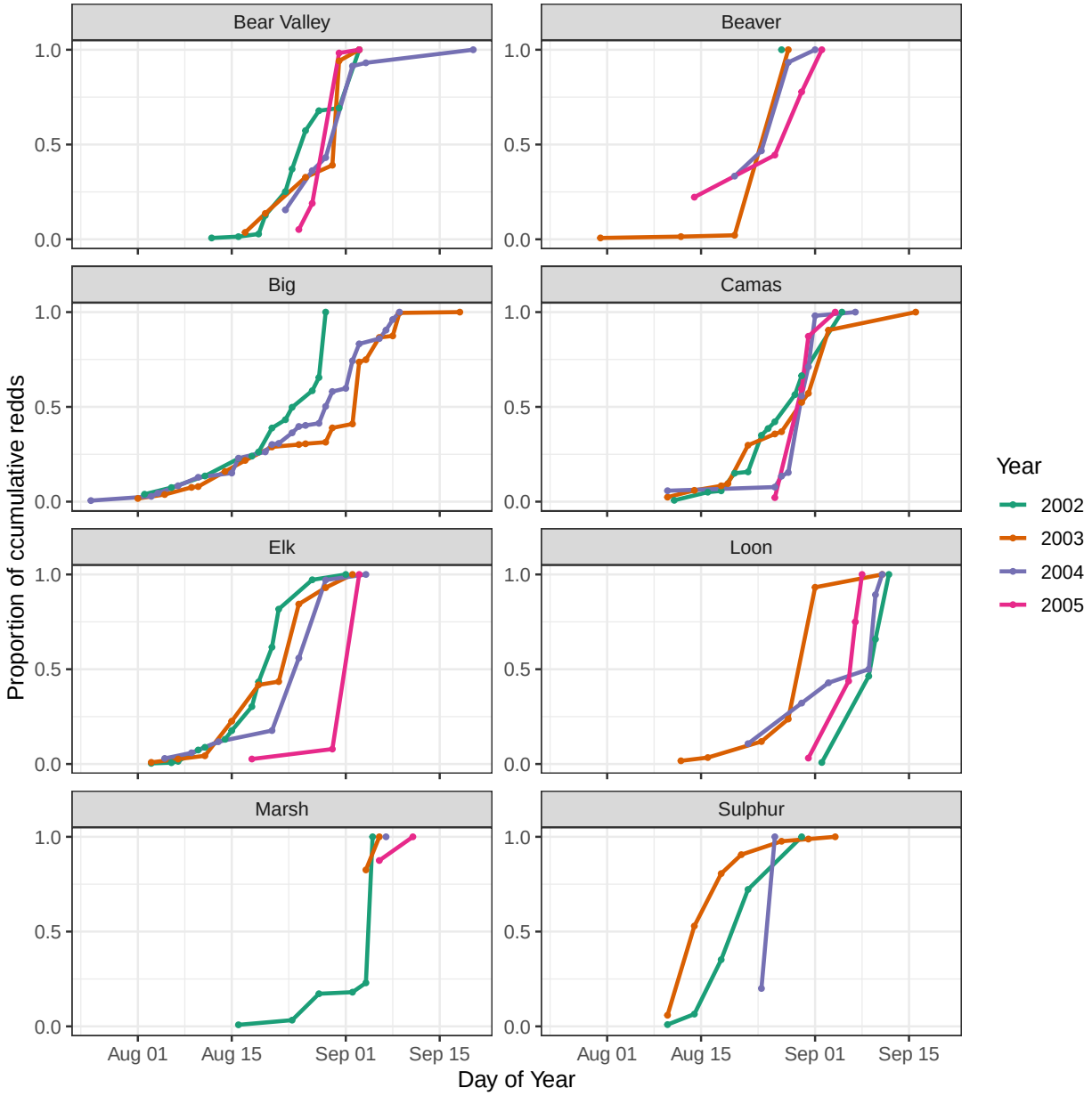


Figure S1.5: Proportion of cumulative Chinook salmon redds over time (day of year) across years (2002–2005) and streams. Each line represents a different year, with color denoting the year. Stream-specific panels illustrate temporal variation in the progression of spawning activity, as measured by cumulative redd counts normalized to the maximum value in each stream-year combination.

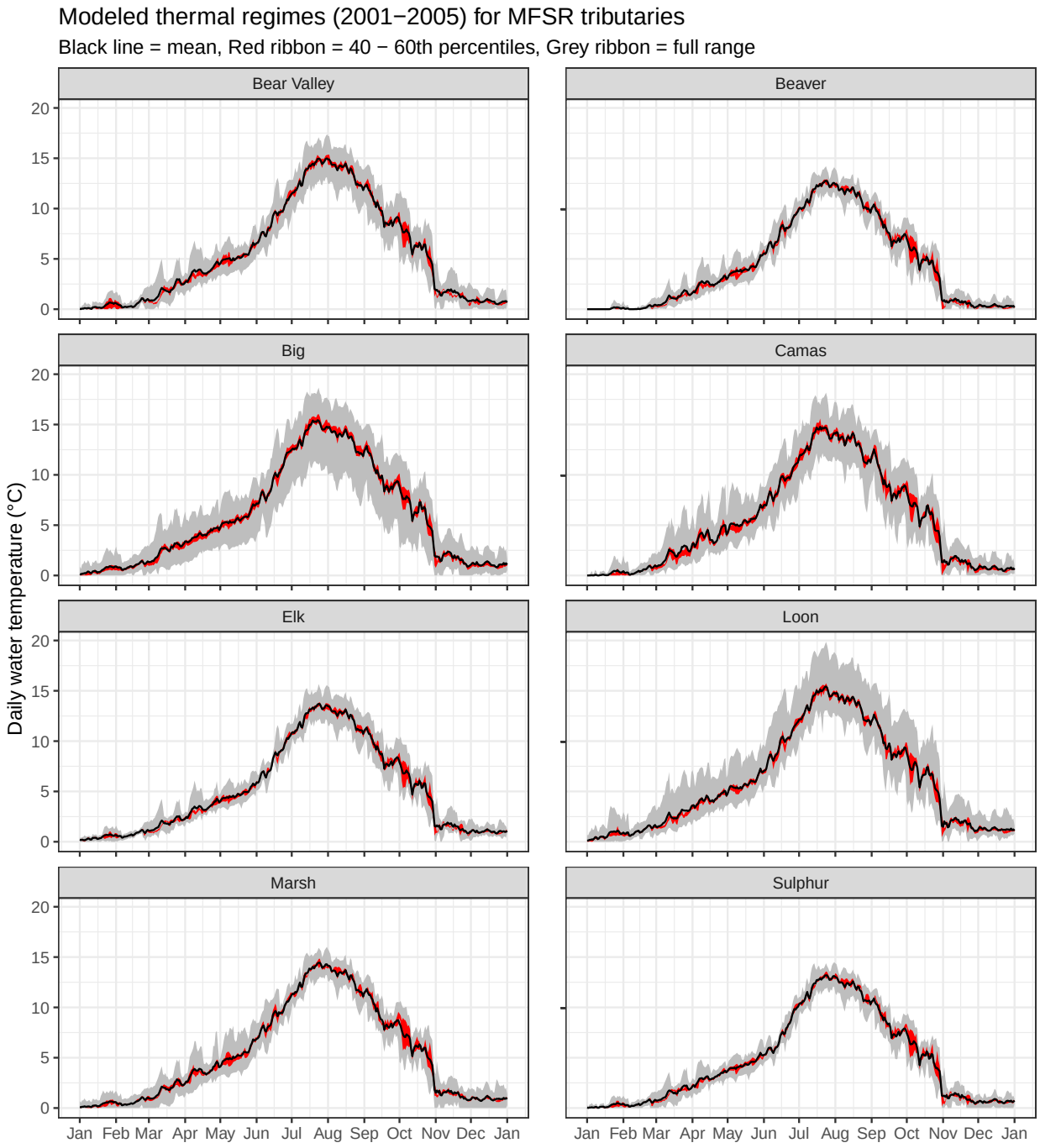


Figure S1.6: Modeled thermal regimes (2001-2005) for MFSR tributaries.

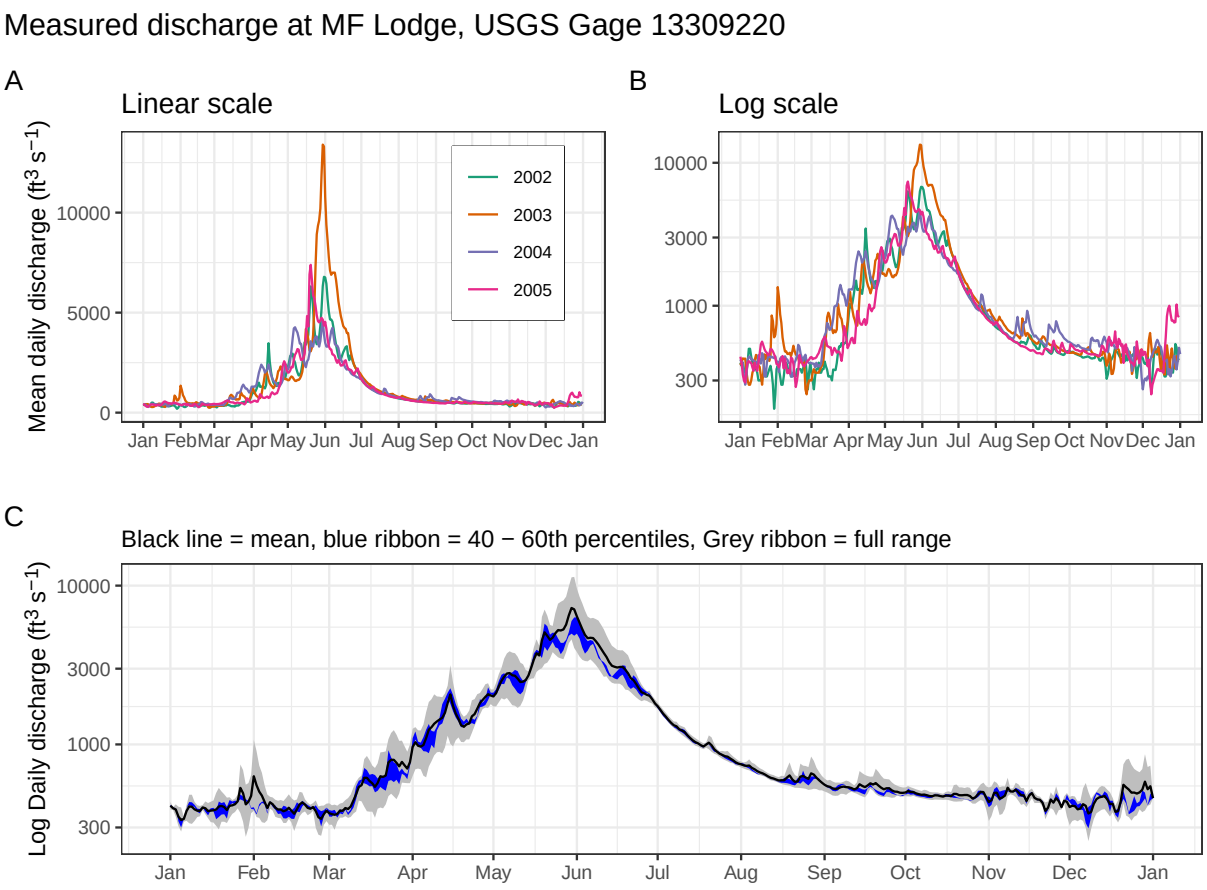


Figure S1.7: Inter-annual variability in daily discharge (cfs) at MF Lodge USGS Gage 13309220.

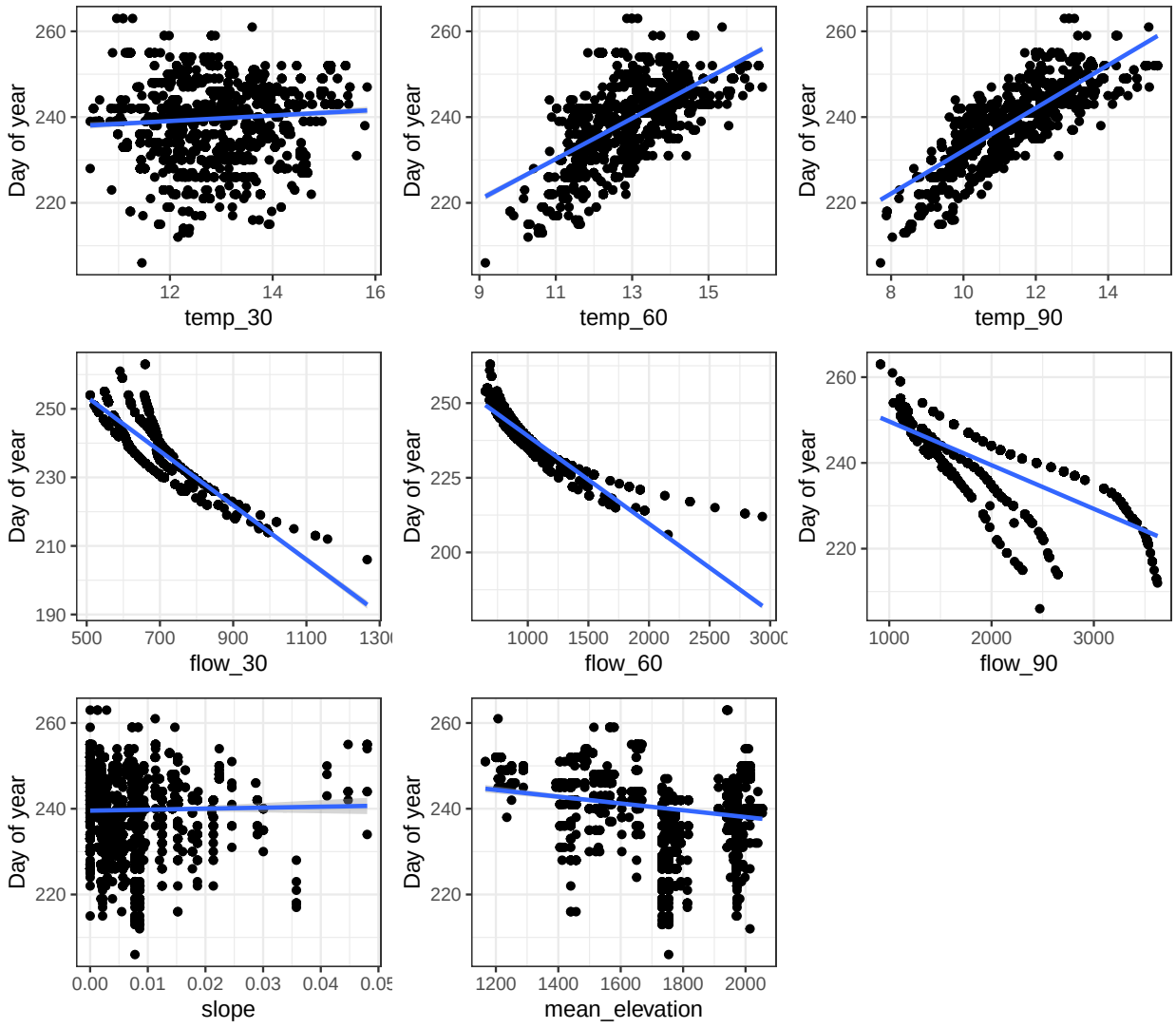


Figure S2.1: Bivariate relationships between spawn date (‘yday‘) and continuous covariates. Solid lines are linear fits.

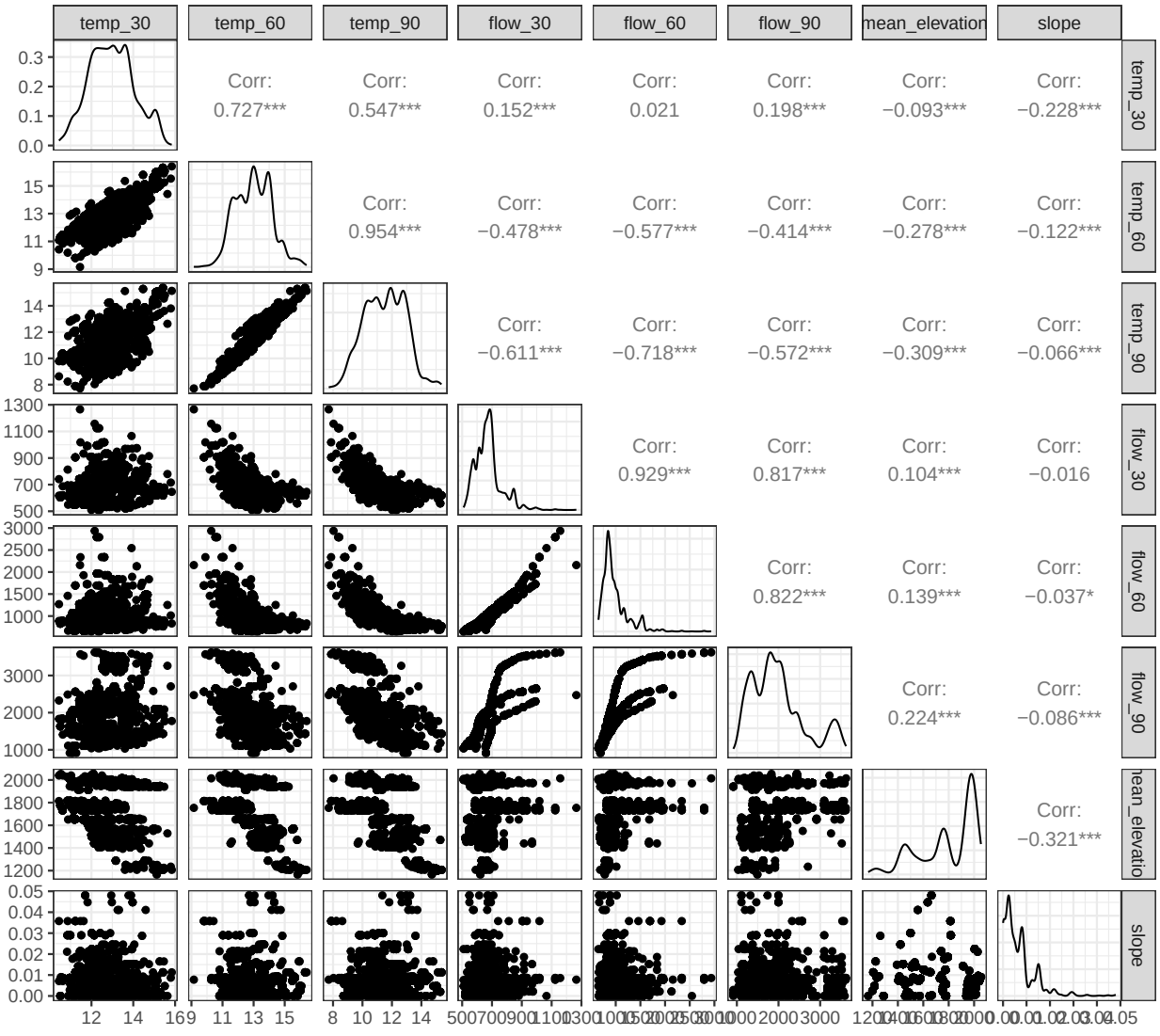


Figure S2.2: Pairwise correlations between continuous covariates.

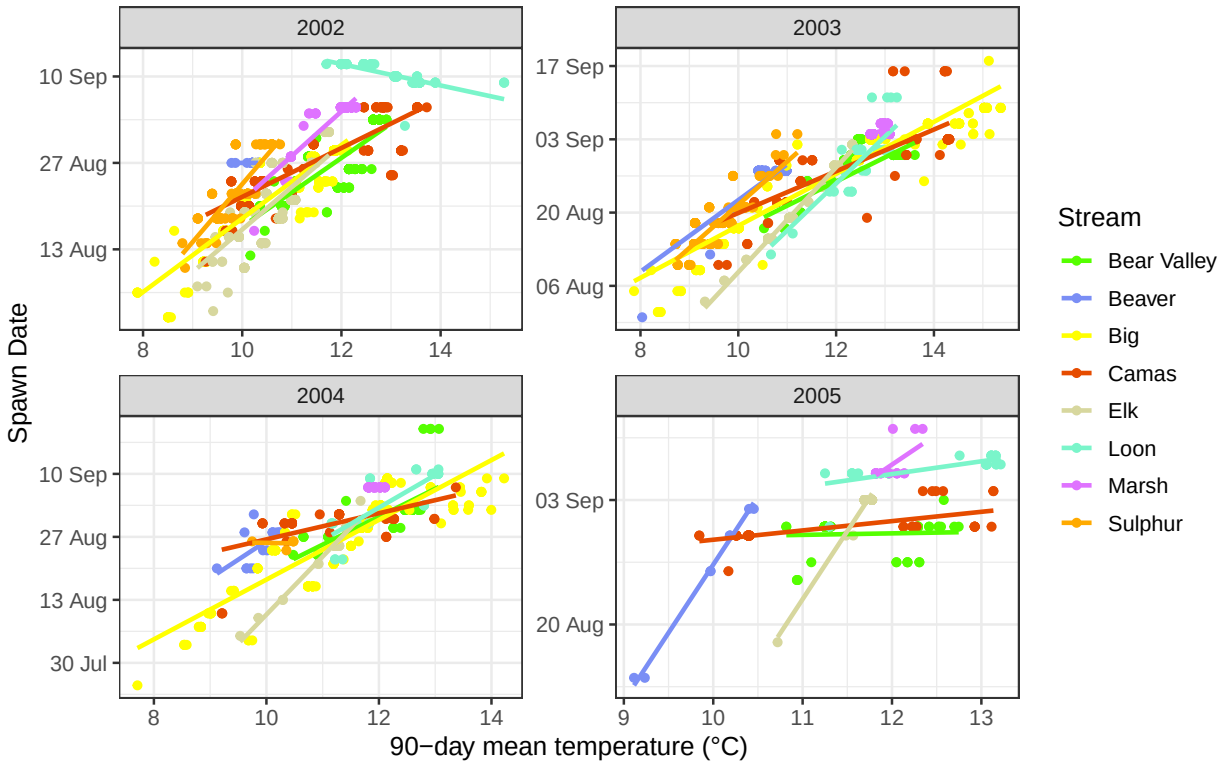


Figure S2.3: Relationship between temp_90 and spawn date by stream and year. Solid lines are linear fits.

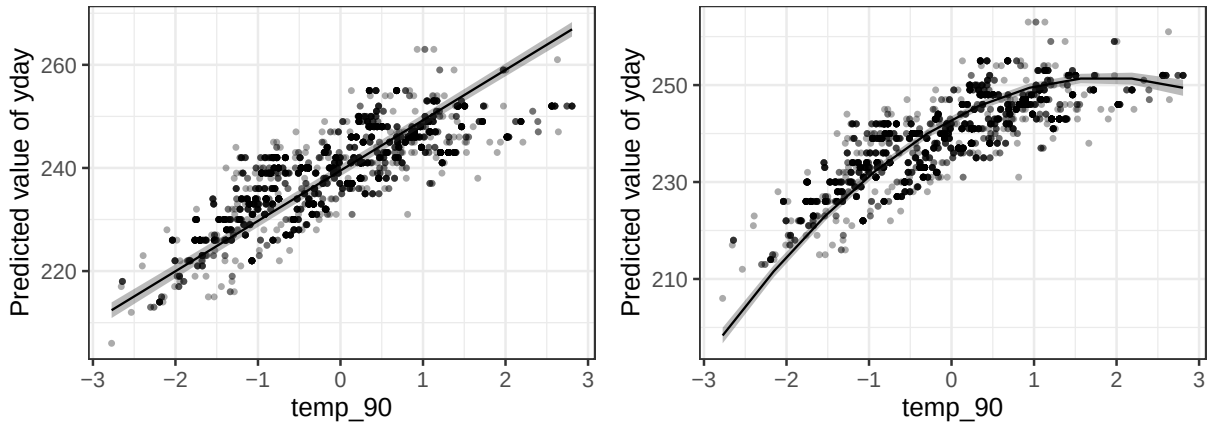


Figure S2.4: Predicted marginal effect of 'temp_90' on 'yday' (spawn date) for (A) 'm.t90.2' and (B) 'm.t90.3'.

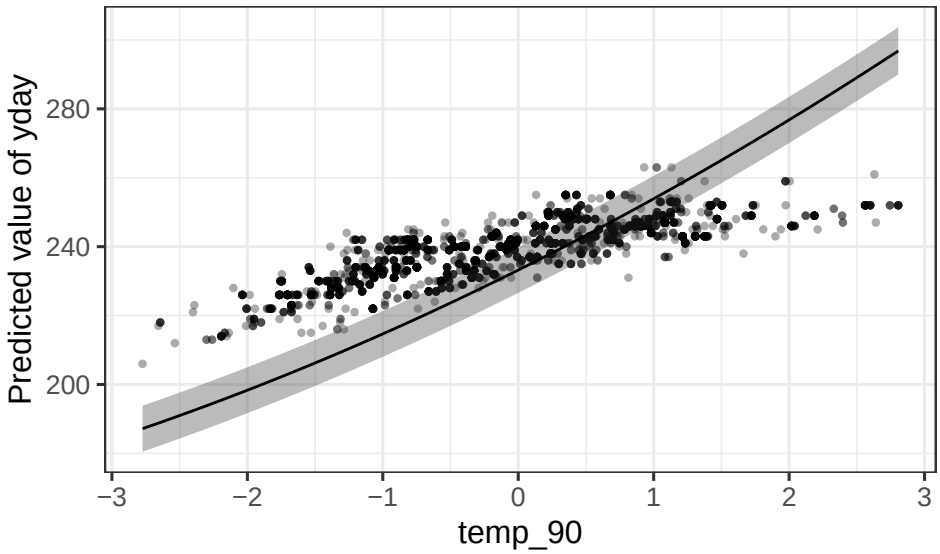


Figure S2.5: Predicted marginal effect (‘m.t90.8‘) of ‘temp_90‘ on ‘yday‘ (accounting for the effects of stream and year).

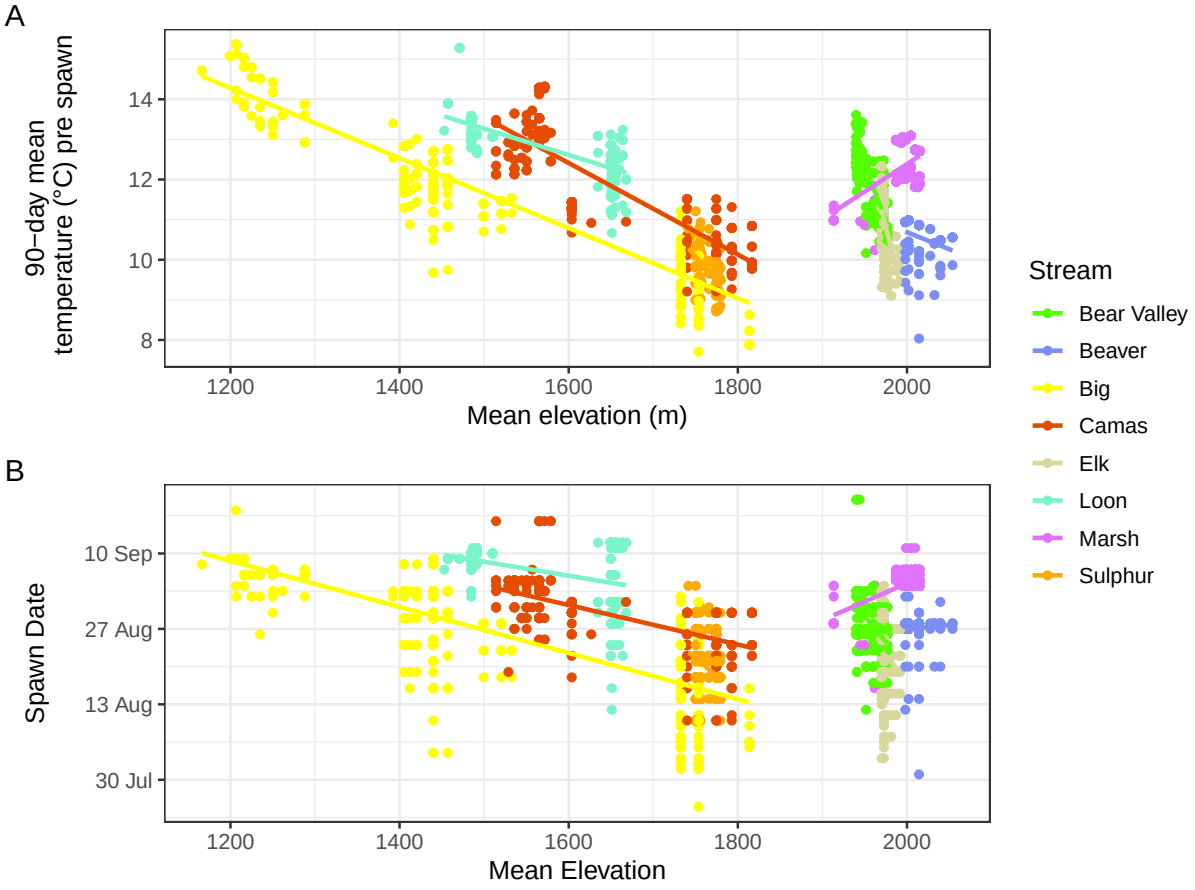


Figure S2.6: Relationship between (A) ‘mean_elevation‘ and ‘temp_90‘ by stream and (B) ‘mean_elevation‘ and ‘yday‘ (spawn date) by stream and year. Solid lines are linear fits.

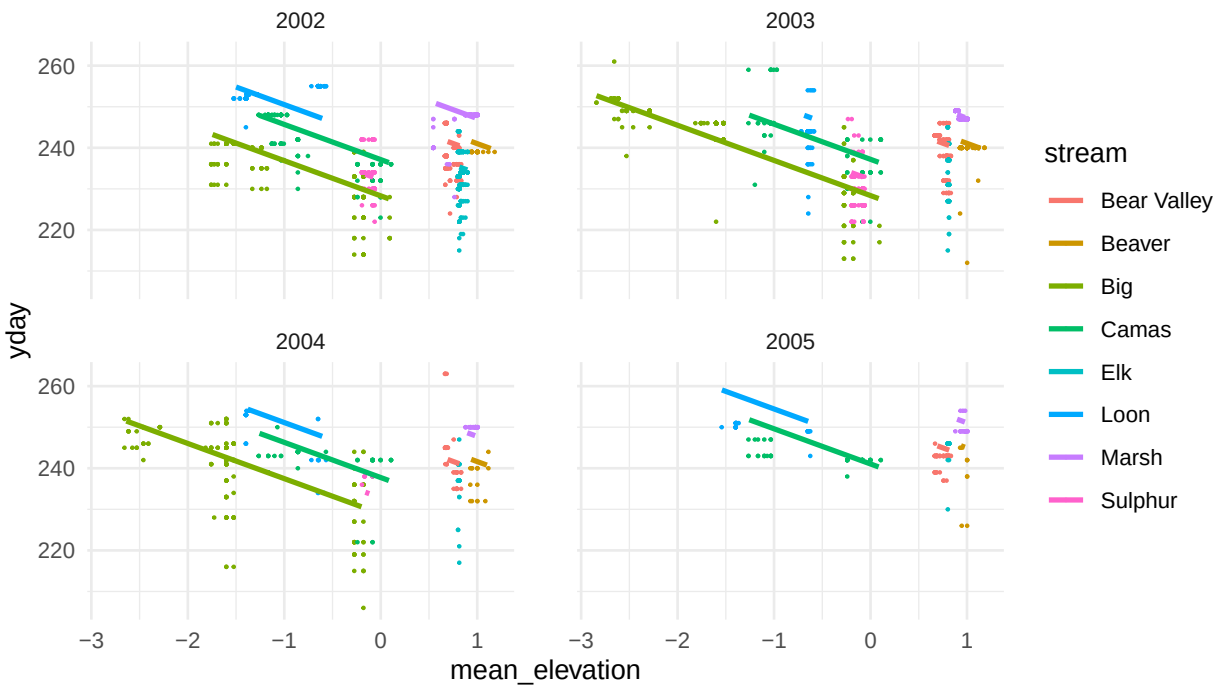


Figure S2.7: Estimated relationship between spawn date and mean_elevation (m.ele.5).

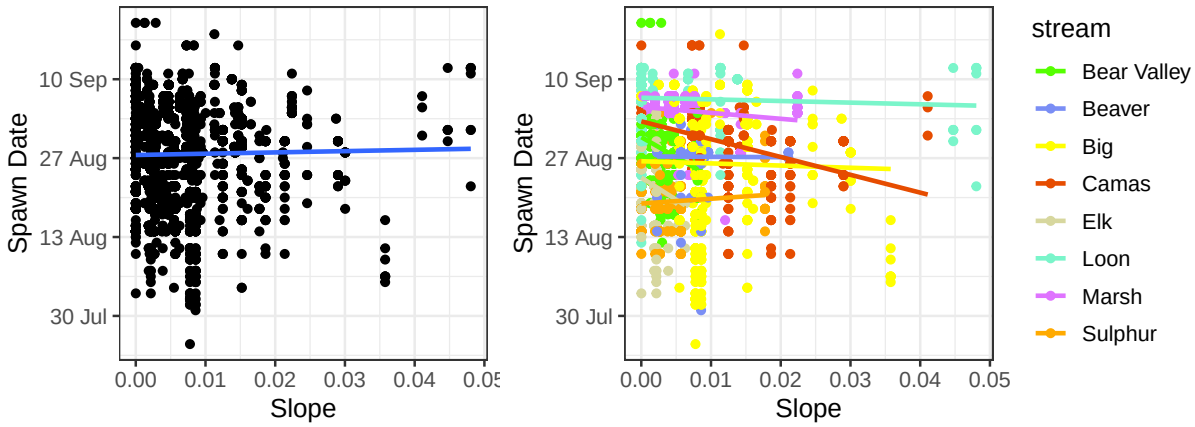


Figure S2.8: Bivariate relationship between 'slope' and spawn date (A) and by stream (B). Solid lines are linear fits.

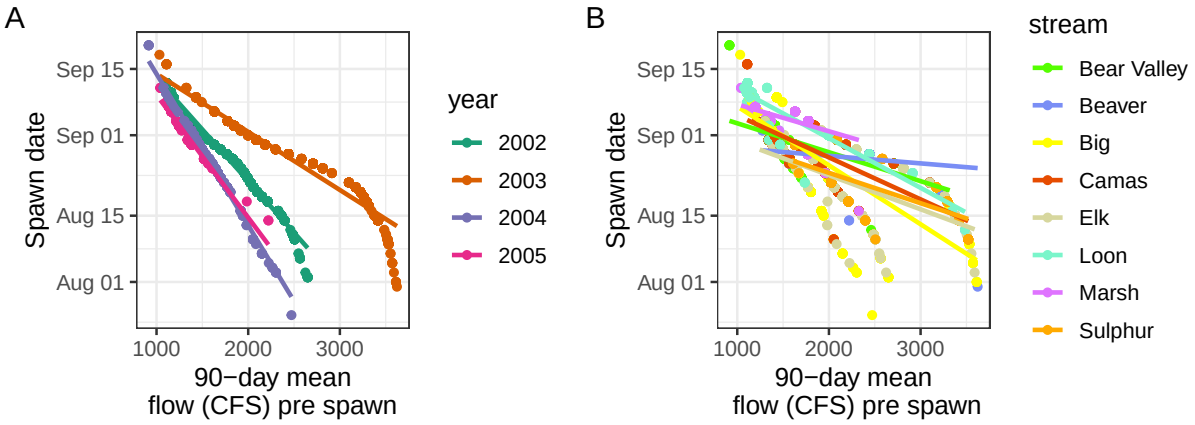


Figure S2.9: Spawn timing vs. 90-day mean discharge pre spawn at MF Lodge.

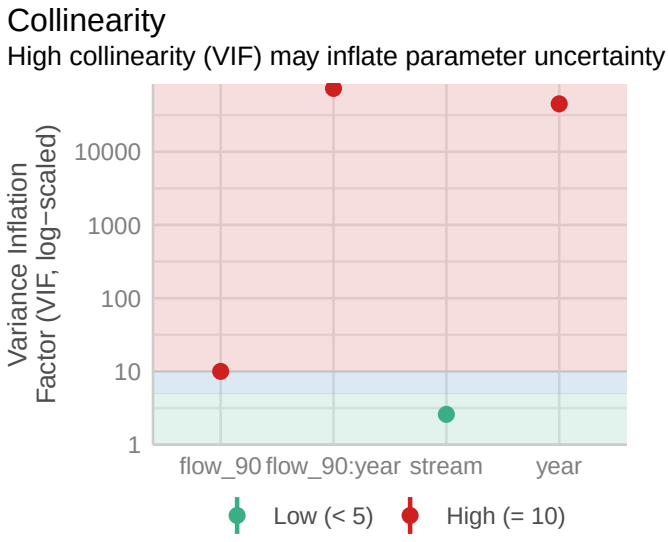


Figure S2.10: Variance inflation factors (VIF) for model 'm7', 'yday flow_90 * year + stream'.

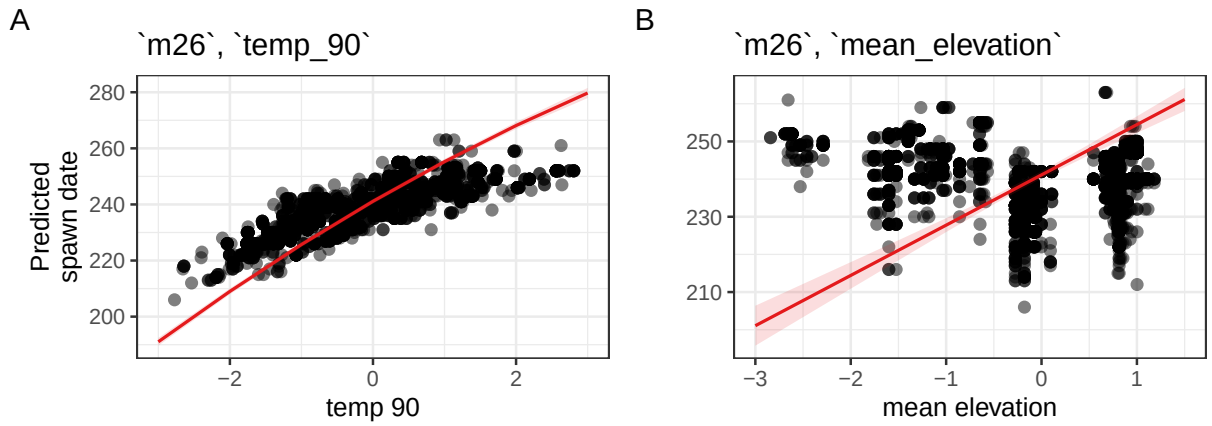


Figure S3.1: Predicted relationship (red line) between spawn date ('yday') and (A) 'temp_90', (B) 'mean_elevation'.

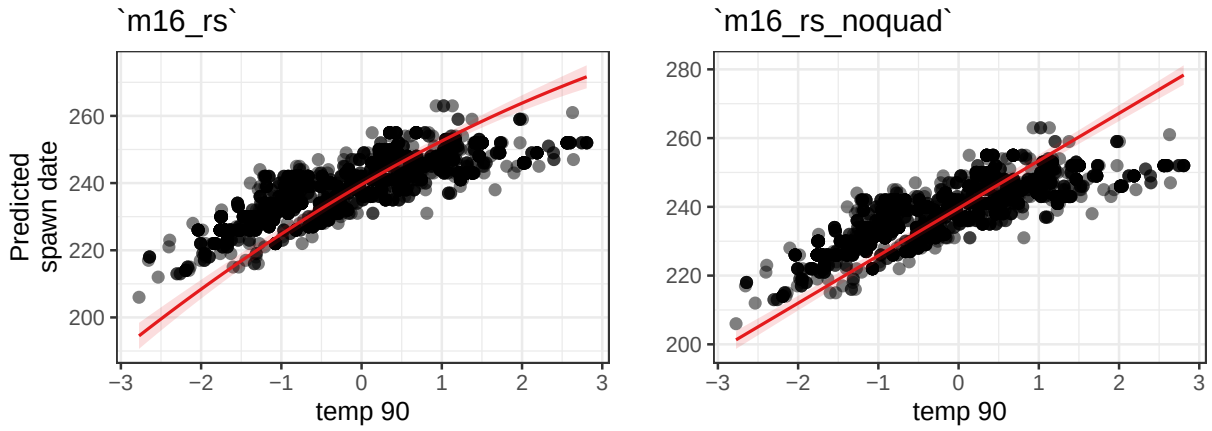


Figure S3.2: Model predictions (red lines) for 'm16_rs' and 'm16_rs_noquad'.

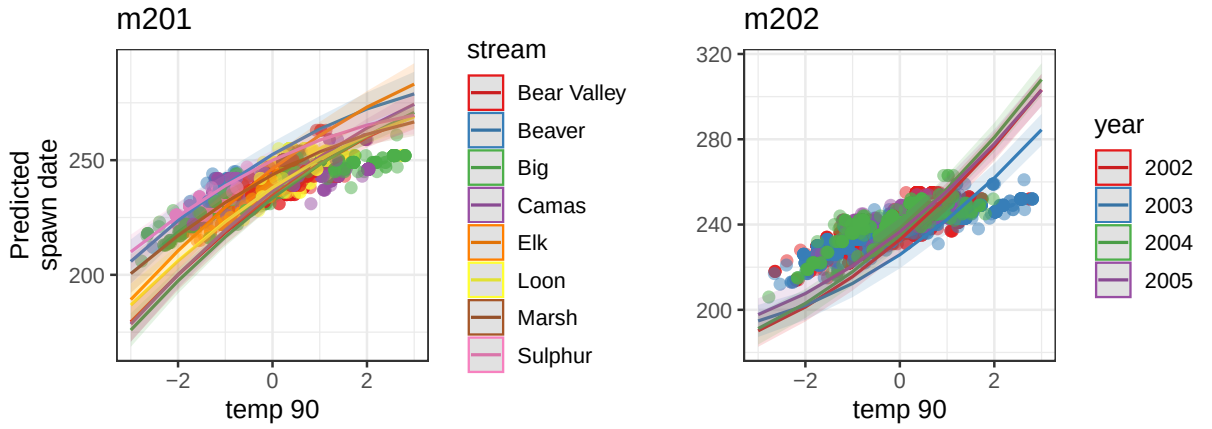


Figure S3.3: Predicted spawn timing.

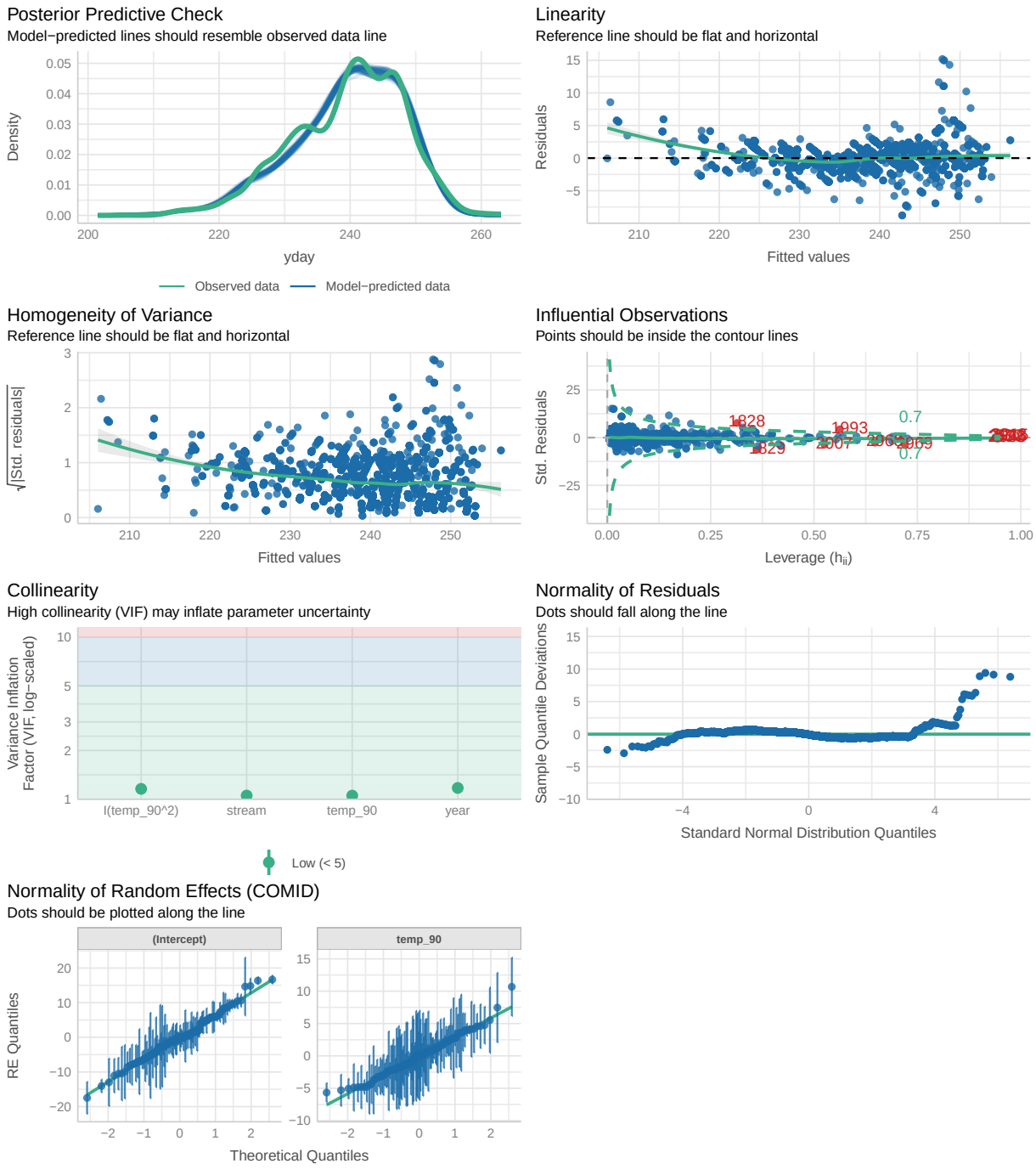


Figure S4.1: Residual diagnostics for ‘mod_final‘.

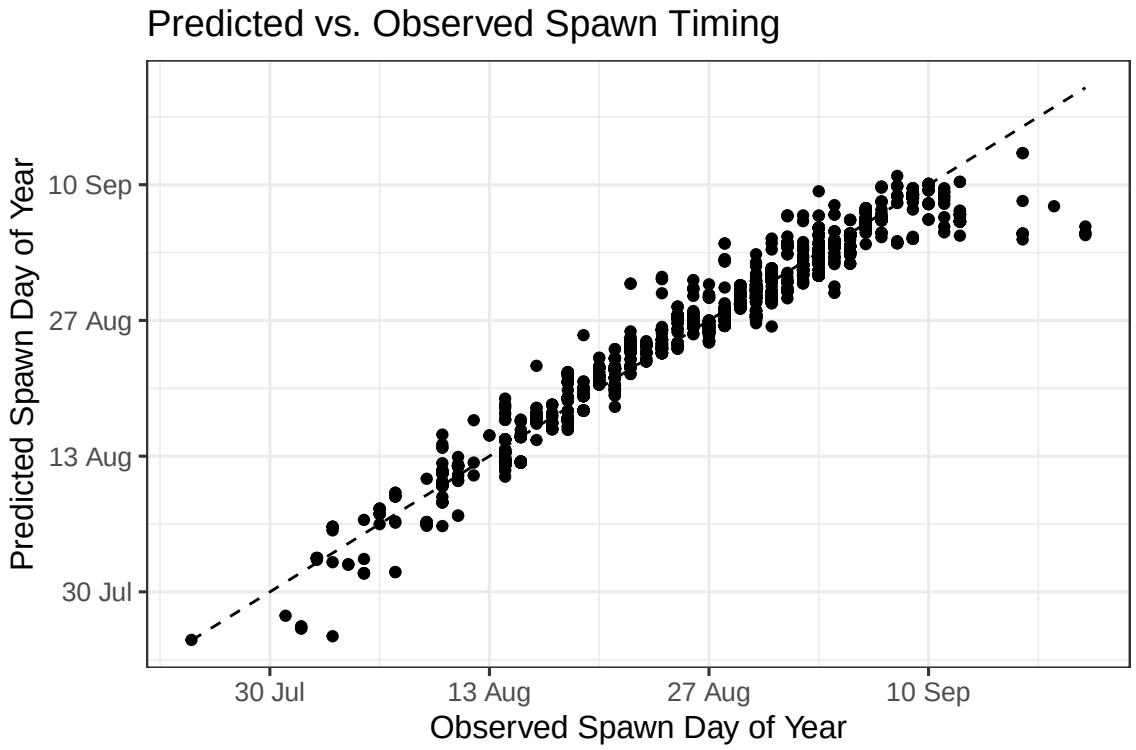


Figure S4.2: Predicted vs. observed spawn timing for Chinook Salmon across all years and streams. The dashed line shows the 1:1 line. Points represent individual redd observations.

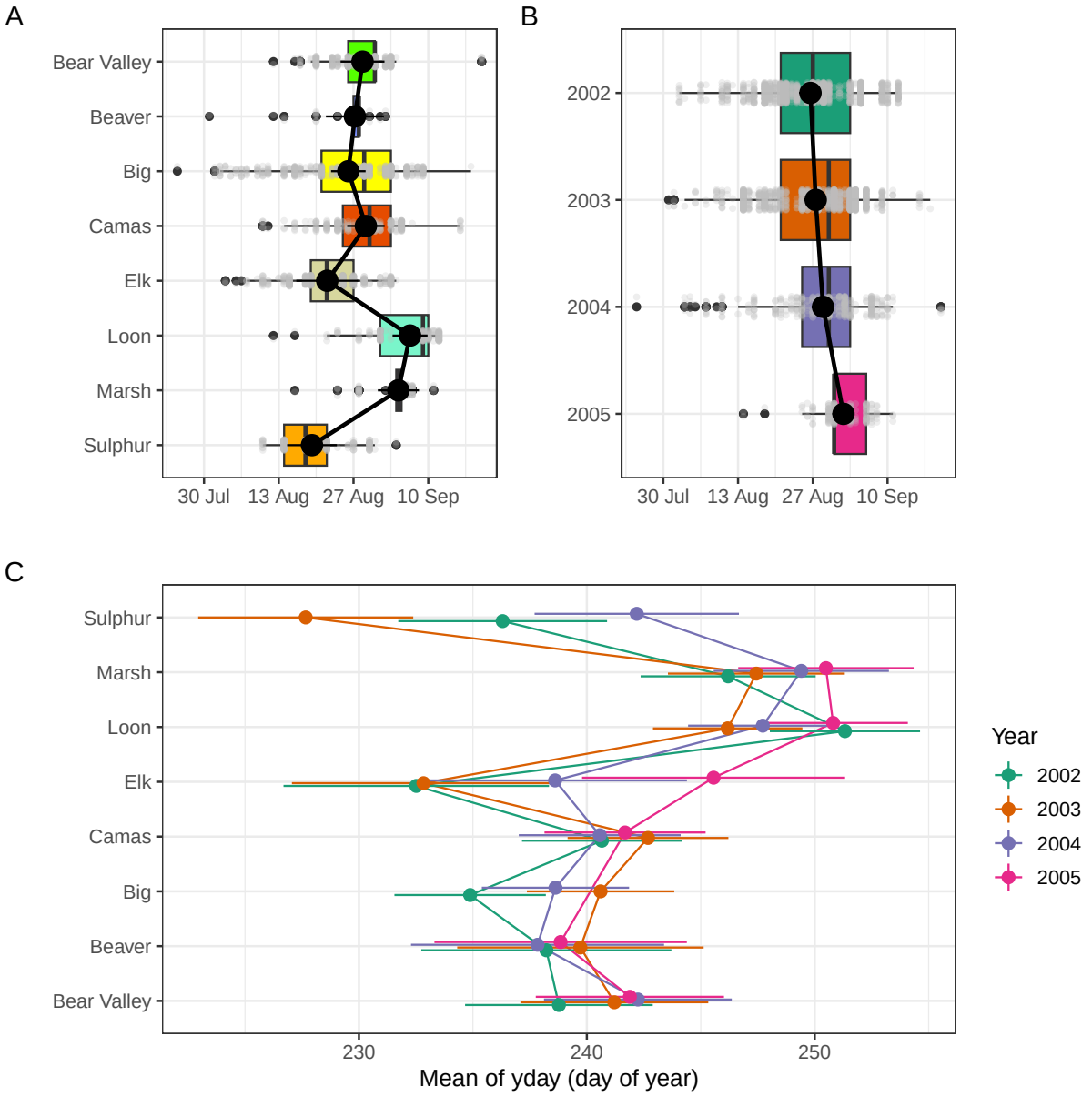


Figure S4.3: Predicted mean spawn dates by stream (A), year (B) and stream and year (C), from the final mixed-effects model. Black points with black lines (A, B), and colored points with horizontal lines (C) represent estimated marginal means and 95% confidence intervals. Boxplots in panels A and B show the distribution of observed redd counts by group, with individual data points in grey. The modeled predictions represent marginal means accounting for fixed effects and averaged over random effects.

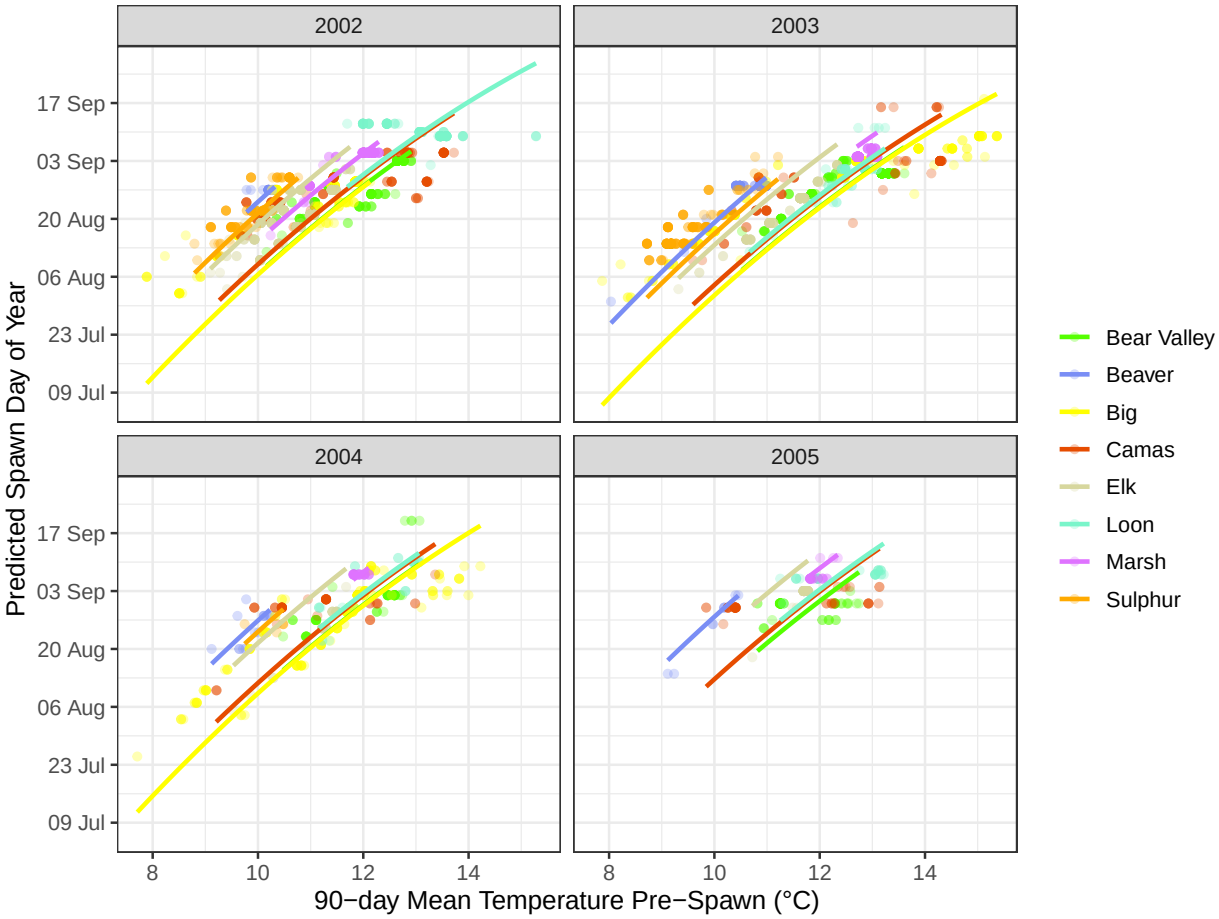


Figure S4.4: Predicted relationship between spawn timing and 90-day pre-spawn mean temperature by stream and year. Lines represent model predictions from the final mixed-effects model. Colored points show observed redd timing, shaded ribbons represent 95% confidence intervals for predictions.

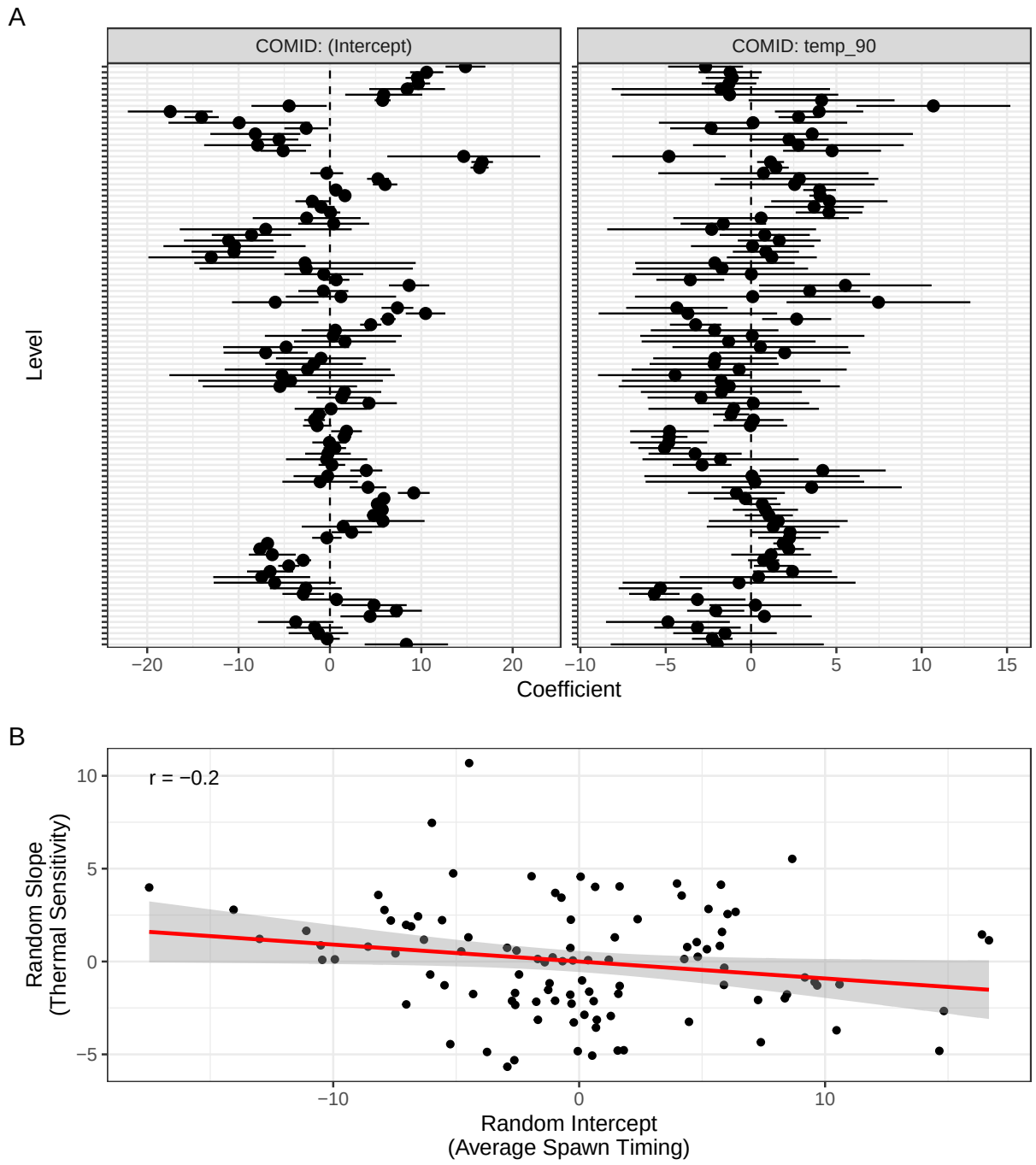


Figure S4.5: (A) COMID-specific random parameter estimates for intercepts (left) and slopes (right). Points represent best linear unbiased predictions (BLUPs) from the final model, with horizontal bars indicating ±1.96 standard errors. (B) Correlation between random intercepts and slopes for 90-day temperature across COMIDs. Each point represents a stream reach (COMID).

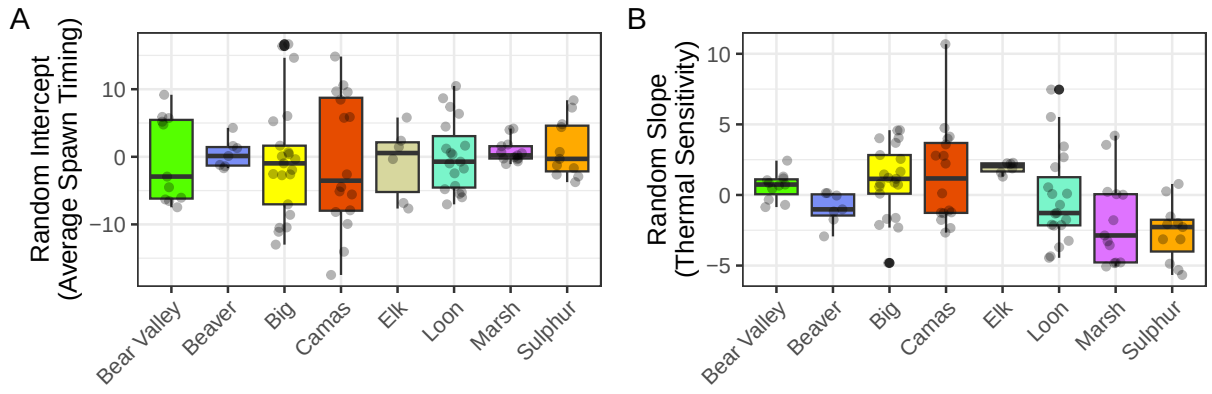


Figure S4.6: Boxplots of random intercepts and slopes for 90-day pre-spawn temperature by stream. Each box represents the distribution of best linear unbiased predictions (BLUPs) for a COMID's random intercept (average spawn timing) or slope (thermal sensitivity). Bars are colored by stream.

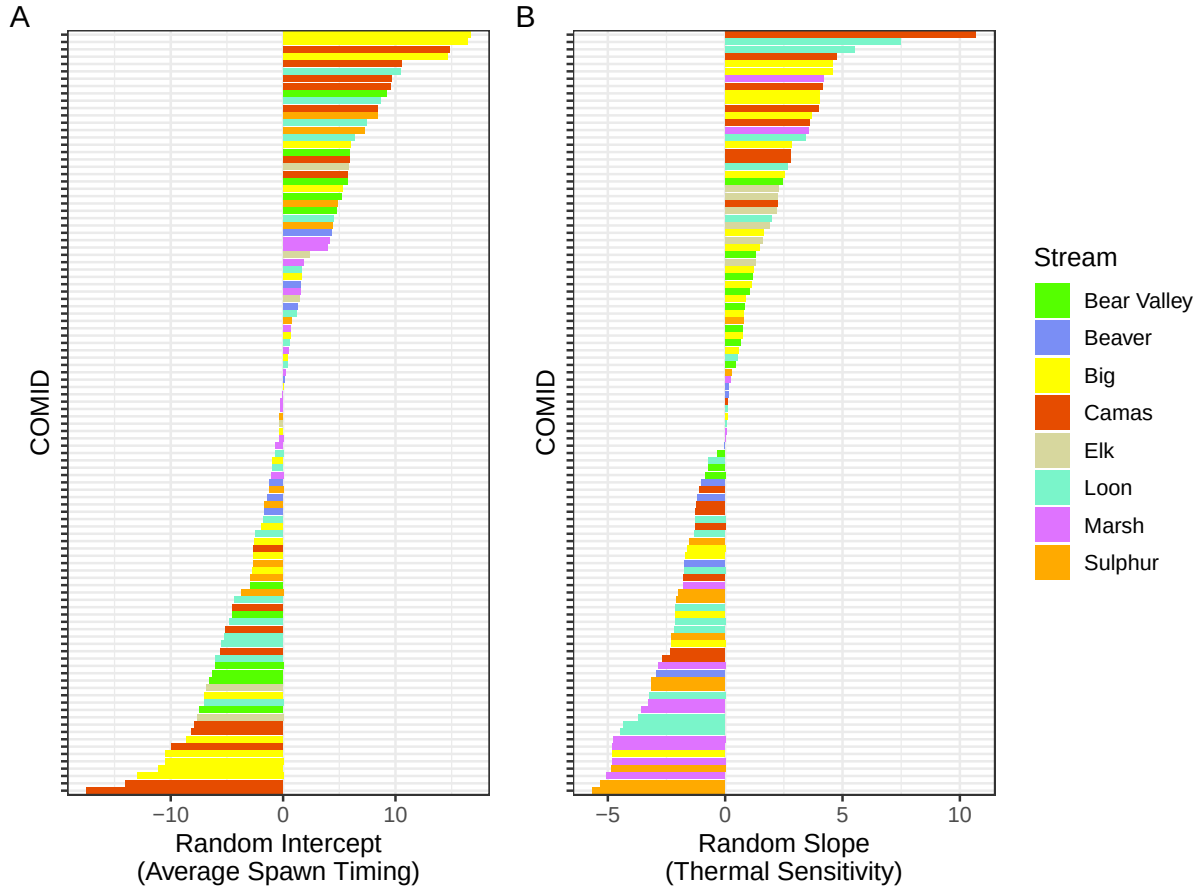


Figure S4.7: Random intercepts and slopes for 90-day pre-spawn temperature by stream. Each bar represents the best linear unbiased prediction (BLUP) for a COMID's random intercept (average spawn timing) or slope (thermal sensitivity). Bars are colored by stream.

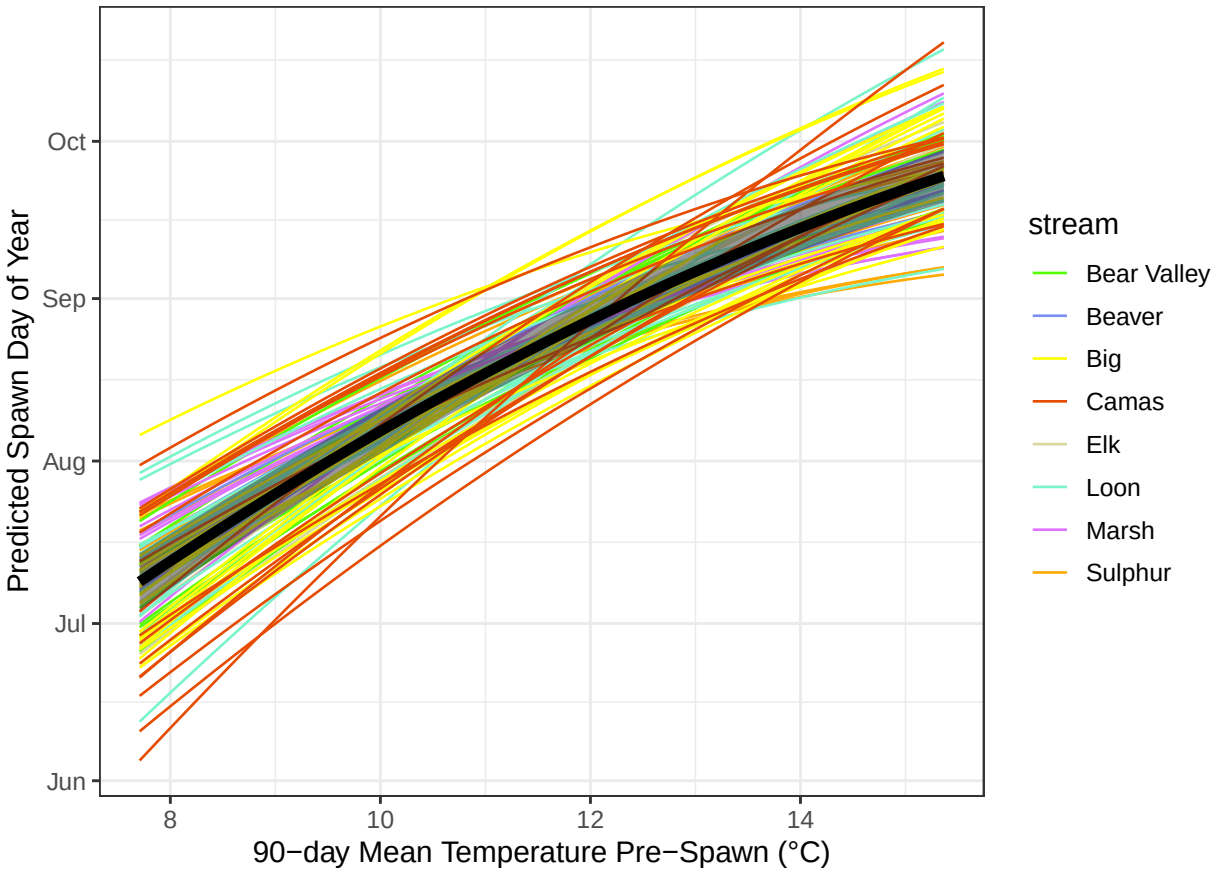


Figure S4.8: Predicted spawn timing by 90-day pre-spawn temperature and COMID. Each line represents the predicted spawn timing for a specific COMID, colored by stream. The black line and shaded ribbon represents the 95% confidence interval for the population-level predictions.

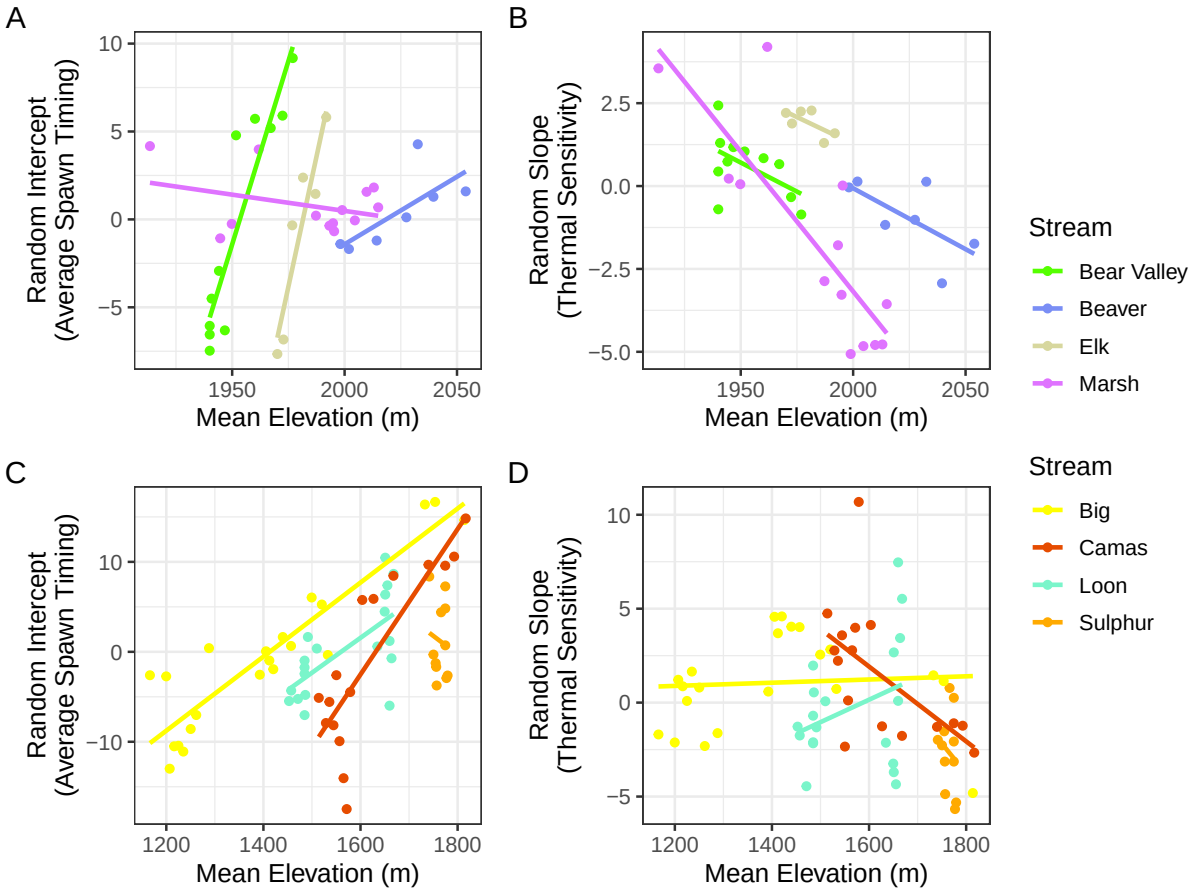


Figure S4.9: Relationship between 'mean_elevation' and (A) random intercepts and (B) slopes for 90-day pre-spawn temperature, as well as 'mean_elevation' and (C) spawn date ('yday') and (D) 'temp_90' . Points and lines are colored by stream, and solid lines represent linear fits.